

Saidha, Tekchand

From: Saidha, Tekchand
Sent: Monday, April 14, 2003 8:17 AM
To: Spencer, Mark
Subject: STIC copy requested - 09/622419

The stic copy for - 09/622419 is requested.

Thanks !

*Jekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595*

STIC-Biotech/ChemLib

9/429

Fr m: Saidha, Tekchand
Sent: Monday, April 14, 2003 9:49 AM
To: STIC-Biotech/ChemLib
Subject: 09/622419 - sequence search request

RECEIVED
APR 14 2003
(STIC)

09/622419

Please search the data base and interference files for :

SEQ ID NO : 1 & 3

Thank you!

Tekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Searcher: _____	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: _____	Bibliographic: _____	DRLink: _____
Date Completed: <u>4/22/03</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: _____	Full text: _____	Sequence Sys.: _____
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:03:55 ; Search time 3213.55 Seconds

(Without alignments)
10459.998 Million cell updates/sec

Title: US-09-622-419-1

Perfect score: 1155
Sequence: 1 atgcgaacacacatttctac.....atgctgcgcgtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_hlg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_in: 18: em_mu: 19: em_om: 20: em_or: 21: em_ov: 22: em_pal: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_un: 28: em_vl: 29: em_vl: 30: em_hlg_hum: 31: em_hlg_inv: 32: em_hlg_other: 33: em_hlg_mus: 34: em_hlg_pln: 35: em_hlg_rod: 36: em_hlg_mam: 37: em_hlg_vrt: 38: em_sy: 39: em_higo_hum: 40: em_higo_mus: 41: em_higo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1155	100.0	1155	6	AX018960	AX018960 Sequence
2	1155	100.0	1155	6	E38344	E38344 Process for
3	1155	100.0	3720	6	AX018968	AX018968 Sequence
4	1155	100.0	4975	6	AX018972	AX018972 Sequence
5	1155	100.0	12354	1	AE000377	AE000377 Escherich
6	1155	100.0	12354	6	AX370209	AX370209 Sequence
7	1155	100.0	12354	6	AX370245	AX370245 Sequence
8	1155	100.0	12354	6	AX370258	AX370258 Sequence
9	1155	100.0	141744	1	ECU28377	ECU28377 Escherich
10	1142.2	98.9	11178	1	AE005524	AE005524 Escherich
11	1142.2	98.9	26658	1	AP002363	AP002363 Escherich
12	1129.4	97.8	1152	1	ECOMETX	M98266 Escherich
13	1043.4	90.3	1462	1	ECOMETX	K02129 E. coli met
14	929.4	80.5	22204	1	AE008842	AE008842 Salmonella
15	927.8	80.3	230050	1	AL627277	AL627277 Salmonella
16	812.6	70.4	10686	1	AE013933	AE013933 Yersinia
17	812.6	70.4	208050	1	AJ414145	AJ414145 Yersinia
18	763.4	66.1	10275	1	AE004133	AE004133 Vibrio ch
19	641.4	55.5	10274	1	U32797	U32797 Haemophilus
20	641.4	55.5	10578	1	AE006143	AE006143 Pasteurel
21	612	53.0	229426	2	AC010532	AC010532 Homo sapi
22	586.2	50.8	10297	1	AE004491	AE004491 Pseudomon
23	560	48.5	12390	1	AE014115	AE014115 Buchnera
24	548.8	47.5	293181	1	AP001119	AP001119 Buchnera
25	533	46.1	10385	1	AE011713	AE011713 Xanthomon
26	526.6	45.6	12556	1	AE012176	AE012176 Xanthomon
27	505	43.7	349061	1	NMA22491	AL162753 Neisseria
28	504.4	43.7	215050	1	AL646057	AL646057 Ralstonia
29	503.4	43.6	10657	1	AE002529	AE002529 Neisseria
30	503.4	43.6	349980	6	AX044034	AX044034 Sequence
31	489.8	42.4	9995	1	AE003891	AE003891 xy1ella f
32	463.4	40.1	1149	6	AX276423	AX276423 Sequence
33	463.4	40.1	27334	6	AX276384	AX276384 Sequence
34	460	39.8	62909	6	AX067457	AX067457 Sequence
35	447.6	38.8	6040	1	RPR238756	AJ238756 Rickettsi
36	447.6	38.6	7235	1	RTY238757	AJ238757 Rickettsi
37	446	38.6	5864	1	HPFMSY	Y11784 R. prowazeki
38	446	38.6	6040	1	RPR238755	AJ238755 Rickettsi
39	446	38.6	237523	1	RPRX04	AJ235273 Rickettsi
40	432.4	37.4	132050	1	ML0672113	AL672113 Mesorhizo
41	422.8	36.6	346820	1	AP003008	AP003008 Mesorhizo
42	418.2	36.2	8033	1	RHE238762	AJ238762 Rickettsi
43	410.8	35.6	9007	1	RFE238763	AJ238763 Rickettsi
44	403.8	35.0	40356	1	SC16	AL159139 Streptomy
45	391.4	33.9	1203	6	AX432557	AX432557 Sequence

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
AX018960	AX018960	Sequence 1 from Patent WO942591.	AX018960	AX018960	GI:10043055	Escherichia coli.	Escherichia coli	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;	Schroeder, H.	Method for producing bloctin	Patent: WO 942591-A 1 26-AUG-1999.
AX018960	AX018960	Sequence 1 from Patent WO942591.	AX018960	AX018960	GI:10043055	Escherichia coli.	Escherichia coli	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;	Schroeder, H.	Method for producing bloctin	Patent: WO 942591-A 1 26-AUG-1999.

FEATURES
Location/Qualifiers
1..1155
/organism="Escherichia coli"
/db_xref="taxon:562"
1..1155
/note="unamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07714.1"
/db_xref="GI:10043056"
/translation="MAKHLFSESSESGHPKINDOISDAVLAILFEDPKARVACFT
YKTKGVLVGEITTSAMVDIEETRTNTRVIGVSHSDMGKDNASCALSLICQSPD
INQVDRADPLEQDQGLMEFGATNETDVLMPATITAHRLVQAEVKKNGTLFW
LRPAKSOVQEDYDQKIVGIDAVLSTOSHSEEDOKSLOEAVVEELIKPLPAEWLT
SATKFFINPTGRFVIGGPMGDCGLTGRIIIVDTYGMAHGHGAFSGKSDSKVRSAA
YAAVYVAKNIIVAAGLADRCIEIOVSAIGVAPEPTIMETEGTEVPEEOLTLVREFP
DLRVEGLIOMDLHPYKEXFAAYGCHREHFEPEKDKMOLLBDAAGLK"

BASE COUNT 277 a 312 c 303 g 263 t
ORIGIN

Query Match 100.0%; Score 1155; DB 6; Length 1155;
Best Local Similarity 100.0%; Pred. No. 1,2e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGCGAAAACACCTTTTACGTCGAGTCGCTCTGTGAAGGCGATCCGACAAATTGCT 60
DB 1 ATGCGAAAACACCTTTTACGTCGAGTCGCTCTGTGAAGGCGATCCGACAAATTGCT 60
OY 61 GACCAAAATTTGATGCGCTTTTAGACGCGATCTCGAAGAGATCCGAAAGCAGCGTT 120
DB 61 GACCAAAATTTGATGCGCTTTTAGACGCGATCTCGAAGAGATCCGAAAGCAGCGTT 120
OY 61 GACCAAAATTTGATGCGCTTTTAGACGCGATCTCGAAGAGATCCGAAAGCAGCGTT 120
DB 61 GACCAAAATTTGATGCGCTTTTAGACGCGATCTCGAAGAGATCCGAAAGCAGCGTT 120
OY 121 GCTTGGCAACCTACGTAAGAACCGGCGATGTTTGTGCGCGGGAATCAGCACAGC 180
DB 121 GCTTGGCAACCTACGTAAGAACCGGCGATGTTTGTGCGCGGGAATCAGCACAGC 180
OY 121 GCTTGGCAACCTACGTAAGAACCGGCGATGTTTGTGCGCGGGAATCAGCACAGC 180
DB 121 GCTTGGCAACCTACGTAAGAACCGGCGATGTTTGTGCGCGGGAATCAGCACAGC 180
OY 181 GCGTGGTAGACATGGAAGAGATCGCGTAACACCGTTCGGGAATTGGCATATGAT 240
DB 181 GCGTGGTAGACATGGAAGAGATCGCGTAACACCGTTCGGGAATTGGCATATGAT 240
OY 241 TCCGACATGGGCTTTAGACGCTCTGTGCGGTTCTGAGCGGCTATCGCAACAGTCT 300
DB 241 TCCGACATGGGCTTTAGACGCTCTGTGCGGTTCTGAGCGGCTATCGCAACAGTCT 300
OY 241 TCCGACATGGGCTTTAGACGCTCTGTGCGGTTCTGAGCGGCTATCGCAACAGTCT 300
DB 241 TCCGACATGGGCTTTAGACGCTCTGTGCGGTTCTGAGCGGCTATCGCAACAGTCT 300
OY 301 CCTGACATCAACACGAGCGGTGACCGTCCGATCCGCTGGAACACGCGGGTGACAG 360
DB 301 CCTGACATCAACACGAGCGGTGACCGTCCGATCCGCTGGAACACGCGGGTGACAG 360
OY 361 GGTGATGTTGGCTAGCGAACTATGAACCGGCGTGTGATCCGACACCTATCAC 420
DB 361 GGTGATGTTGGCTAGCGAACTATGAACCGGCGTGTGATCCGACACCTATCAC 420
OY 421 TATGACACCGCTCTGTACAGCGCTGAGTGCCTAAACACGCACTTGCCTGG 480
DB 421 TATGACACCGCTCTGTACAGCGCTGAGTGCCTAAACACGCACTTGCCTGG 480
OY 481 CTGCGCCCGGACGCAAAAGCCAGTGTTCAGTATGACGACGCGCAAAATCGTTGCT 540
DB 481 CTGCGCCCGGACGCAAAAGCCAGTGTTCAGTATGACGACGCGCAAAATCGTTGCT 540
OY 541 ATCGATGCTGTGCTCTTCCACCTACGACCTTGAAGAGATGACACAGAAATCGCTGCA 600
DB 541 ATCGATGCTGTGCTCTTCCACCTACGACCTTGAAGAGATGACACAGAAATCGCTGCA 600
OY 601 GAAGCGTATGAGAGATCATCAAGCAATTCGCCCGCTGATGCTGCTCTGCG 660
DB 601 GAAGCGTATGAGAGATCATCAAGCAATTCGCCCGCTGATGCTGCTCTGCG 660
OY 661 ACCAAATTTCTCATCAACCGGCGGCTGTTGTTATCGGTGCGCAATGGGTACTGC 720
DB 661 ACCAAATTTCTCATCAACCGGCGGCTGTTGTTATCGGTGCGCAATGGGTACTGC 720
```

```

OY 721 GGTCTGACTGTCGTAATAATATGTTGATACCTACGGCGGCATGCGCGCTCAGCGTCC 780
DB 721 GGTCTGACTGTCGTAATAATATGTTGATACCTACGGCGGCATGCGCGCTCAGCGTCC 780
OY 781 GGTGATCTCTCTGTGAAGATCATCAAAAGTGAACCGCTTCGAGAGCTTACGACAGCT 840
DB 781 GGTGATCTCTCTGTGAAGATCATCAAAAGTGAACCGCTTCGAGAGCTTACGACAGCT 840
OY 841 TATGTCGGGAAACATCGTTGCTGCTGCGCTGCGCCGATGCTTGAATTCACGTTTCC 900
DB 841 TATGTCGGGAAACATCGTTGCTGCTGCGCTGCGCCGATGCTTGAATTCACGTTTCC 900
OY 901 TACGCAATCGGCGTGGCTGGAACCGACCTCCATCATGATGAGAACTTTGCTGATGACAAA 960
DB 901 TACGCAATCGGCGTGGCTGGAACCGACCTCCATCATGATGAGAACTTTGCTGATGACAAA 960
OY 961 GTGCTTCTGAAACACTGACCTCTGCTGATGATGTTCTTGACCTGCGCCCATATGCT 1020
DB 961 GTGCTTCTGAAACACTGACCTCTGCTGATGATGTTCTTGACCTGCGCCCATATGCT 1020
OY 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCAAAAGAACCGGATACGCTAC 1080
DB 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCAAAAGAACCGGATACGCTAC 1080
OY 1081 TTTGCTGCTGAAACATTTCCCGTGGGAAAAACCGAACGCGAGCTGCGGATGCT 1140
DB 1081 TTTGCTGCTGAAACATTTCCCGTGGGAAAAACCGAACGCGAGCTGCGGATGCT 1140
OY 1141 GCCGCTGTGAAGTAA 1155
DB 1141 GCCGCTGTGAAGTAA 1155
```

RESULT 2
E38344
LOCUS E38344 1155 bp DNA linear PAT 31-JAN-2002
DEFINITION Process for producing L-methionine by fermentation.
ACCESSION E38344
VERSION E38344.1 GI:18624956
KEYWORDS JP 2000139471-A/17.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 1155)
AUTHORS Usuta,Y. and Kurahashi,O.
TITLE Process for producing L-methionine by fermentation
JOURNAL Patent: JP 2000139471-A 17 23-MAY-2000;
AJINOMOTO CO INC
OS Escherichia coli
PN JP 2000139471-A/17
PD 23-MAY-2000
PF 17-NOV-1998 JP 1998326717

COMMENT
PR YOSHIHIRO USUTA,OSAMU KURAHASHI
PI PC C12N15/09,C12N1/21,C12N9/04,C12N9/10,C12N9/12,C12N9/88, PC
C12P13/12/15
PC (C12N15/09,C12R1:19),(C12N1/21,C12R1:19),(C12P13/12,C12R1:19),
PC C12N15/00,
PC (C12N15/00,C12R1:19)
CC
FH

FEATURES
source 1..1155 Location/Qualifiers
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 277 a 312 c 303 g 263 t
ORIGIN

Query Match 100.0%; Score 1155; DB 6; Length 1155;
Best Local Similarity 100.0%; Pred. No. 1,2e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

OY 1 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 60
    |||
    1 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 60
OY 61 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 120
    |||
    61 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 120
OY 121 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 180
    |||
    121 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 180
OY 181 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTCCGCAATTTGGCTATGTCAT 240
    |||
    181 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTCCGCAATTTGGCTATGTCAT 240
OY 241 TCCGACATGGGCTTGAGCTAACCTCCCTGCGGTCTGAGCGCTATCCGCAACAGCTCT 300
    |||
    241 TCCGACATGGGCTTGAGCTAACCTCCCTGCGGTCTGAGCGCTATCCGCAACAGCTCT 300
OY 301 CCTGACATCAACACGAGCGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 360
    |||
    301 CCTGACATCAACACGAGCGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 360
OY 361 GCTCTGATGTTTGGCTACGCAACTAATGAACCGAGCTGCTGATGCCACACTATACAC 420
    |||
    361 GCTCTGATGTTTGGCTACGCAACTAATGAACCGAGCTGCTGATGCCACACTATACAC 420
OY 421 TATGCACACCGCTTGGTACAGCGTCAAGCTGAAGTGGTAAACCGGCACTGCCCTGG 480
    |||
    421 TATGCACACCGCTTGGTACAGCGTCAAGCTGAAGTGGTAAACCGGCACTGCCCTGG 480
OY 481 CTGGCGCCGAGCGCAAAAGCCAGGATGACTTTTCACTATGACGAGCGCAAAATTCGTTGGT 540
    |||
    481 CTGGCGCCGAGCGCAAAAGCCAGGATGACTTTTCACTATGACGAGCGCAAAATTCGTTGGT 540
OY 541 ATGCATGCTGTGCTGCTTCCACTCAGCACTCTGAAGAGATCGACAGAAATTCGCTGCAA 600
    |||
    541 ATGCATGCTGTGCTGCTTCCACTCAGCACTCTGAAGAGATCGACAGAAATTCGCTGCAA 600
OY 601 GAAGCGGTATGGAAGAGATCATCAAGCCCAATTTCCCGCTGGAATGGCTGACTTCTGCC 660
    |||
    601 GAAGCGGTATGGAAGAGATCATCAAGCCCAATTTCCCGCTGGAATGGCTGACTTCTGCC 660
OY 661 ACCAAATTTCTCATCAACCCGACCGGCTGTTGTTATGCGTGGCCCAATGGGTGACTGC 720
    |||
    661 ACCAAATTTCTCATCAACCCGACCGGCTGTTGTTATGCGTGGCCCAATGGGTGACTGC 720
OY 721 GGTCTGACTGTCGTAAATTAATCGTTGATACCTACGGCGGATGGCGGTCACGGTGGC 780
    |||
    721 GGTCTGACTGTCGTAAATTAATCGTTGATACCTACGGCGGATGGCGGTCACGGTGGC 780
OY 781 GGTGATTTCTCTGCTAAAGATCCATCAAAAGTGGACCGTTCGGCAGCTTACGACGACGT 840
    |||
    781 GGTGATTTCTCTGCTAAAGATCCATCAAAAGTGGACCGTTCGGCAGCTTACGACGACGT 840
OY 841 TATGTCGCAAAACATCGTCTGCTGCTGCGCGATCGTTGGAATTTAGGTTTCC 900
    |||
    841 TATGTCGCAAAACATCGTCTGCTGCTGCGCGATCGTTGGAATTTAGGTTTCC 900
OY 841 TATGTCGCAAAACATCGTCTGCTGCTGCGCGATCGTTGGAATTTAGGTTTCC 900
    |||
    841 TATGTCGCAAAACATCGTCTGCTGCTGCGCGATCGTTGGAATTTAGGTTTCC 900
OY 901 TACGCAATCGGCGTGGCTGAACCGACCTTCATCATGTTGAAGAACTTTGCGTACTGAGAAA 960
    |||
    901 TACGCAATCGGCGTGGCTGAACCGACCTTCATCATGTTGAAGAACTTTGCGTACTGAGAAA 960
OY 961 GTGCTTCTGAAACCACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
    |||
    961 GTGCTTCTGAAACCACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
OY 1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
    |||
    1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080

```

```

OY 1081 TTTGCTGTAACATTTCCCGTGGGAAATAACCGACAAAGCCAGCTGCTGGCGATGCT 1140
    |||
    1081 TTTGCTGTAACATTTCCCGTGGGAAATAACCGACAAAGCCAGCTGCTGGCGATGCT 1140
OY 1141 GCCGCTGTAAGTAA 1155
    |||
    1141 GCCGCTGTAAGTAA 1155
OY 1141 GCCGCTGTAAGTAA 1155
    |||
    1141 GCCGCTGTAAGTAA 1155

RESULT 3
AX018968
LOCUS AX018968 3720 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 9 from Patent WO942591.
ACCESSION AX018968
VERSION AX018968.1 GI:10043063
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia coli.
REFERENCE 1 (bases 1 to 3720)
AUTHORS Schroeder, H.
TITLE Method for producing biotin
JOURNAL Patent: WO 942591-A 9 26-AUG-1999;
SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES
source location/Qualifiers
1..3720
/organism="Escherichia coli"
/db_xref="taxon:562"
530..1684
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07718.1"
/db_xref="GI:10043064"
/translation="MAKHLFTSEVSEGHPEKIDQISDAVLDA1LEODPARVACET
YVKTGMVLVGEIETTSAYVDIEITRNVRBIGVYVSDMGDANSAYLSA1IGQSPD
INOCVDRADPLEOGADQGLMEFYATNTDVLMPAPITVYARLVORQREKNTLPM
LRPAKSOVTFQYDDGKIVGIDAVLSTQHSSEIDOKSLQAVVMEI1KPLPEMLT
SATKFEINPTGRTFVYGGPMGCGGLGRK1IYDTDGMKRGHGAFFSGKVDPSAA
YVARYVAKNIVIAACIADRCLEIOVSATGVAAPRTSIMETPECTEVPKQLTLVREF
DLRYGLTOMLDLPHITKETAIGHFQREHFPWEKTDKADLRUAGLK"
BASE COUNT 932 a 1019 c 954 g 815 t
ORIGIN

Query Match 100.0% Score 1155: DB 6: Length 3720;
Best Local Similarity 100.0%: Pred. No. 1.2e-266;
Matches 1155: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 60
    |||
    1 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 60
OY 530 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 589
    |||
    530 ATGGCAAAACACCTTTTACGTCGAGATCCGCTCTGAAAGGCGATCTGACAAATTCGT 589
OY 61 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 120
    |||
    61 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 120
OY 590 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 649
    |||
    590 GACCAATTTCTGATGCGCTTTTAGCGGATCCTCGAAGAGATCCGAAGACCGCTT 649
OY 121 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 180
    |||
    121 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 180
OY 650 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 709
    |||
    650 GCTGGCAACCTACGTAATAAACCGGATGTTTATGTTGGCGGCAATTCACACGACG 709
OY 181 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTTCGGAATTTGCTATGTCAT 240
    |||
    181 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTTCGGAATTTGCTATGTCAT 240
OY 710 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTTCGGAATTTGCTATGTCAT 769
    |||
    710 GCCTGGGTAAGACATCGAAGAGATCACCCTTAACACCGTTTCGGAATTTGCTATGTCAT 769
OY 241 TCCGACATGGGCTTTGAGCGTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
    |||
    241 TCCGACATGGGCTTTGAGCGTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
OY 770 TCCGACATGGGCTTTGAGCGTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
    |||
    770 TCCGACATGGGCTTTGAGCGTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
OY 301 CCTGACATCAACACGAGCGGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 360
    |||
    301 CCTGACATCAACACGAGCGGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 360
OY 830 CCTGACATCAACACGAGCGGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 889
    |||
    830 CCTGACATCAACACGAGCGGCTTGACCGTCCGATCCGCTGGAACAGGCGCGGTGACACAG 889

```

```

Oy 361 GGTCTGATGTTGGCTACGCACTAATGAACCGAGCTGGTGAATGCCAGACCTATCACC 420
Db 890 GGTCTGATGTTGGCTACGCACTAATGAACCGAGCTGGTGAATGCCAGACCTATCACC 949
Oy 421 TATGACACCGCTCTGTACAGCGCTCAGGCTGAAGTGGCTAAAAAGGCACTCTGCCGTGG 480
Db 950 TATGACACCGCTCTGTACAGCGCTCAGGCTGAAGTGGCTAAAAAGGCACTCTGCCGTGG 1009
Oy 481 CTGGCCCGGAGCGGAAAGCGAGCTGCTTTTCAGTATGACGAGCGCAAAATCTGTTGGT 540
Db 1010 CTGGCCCGGAGCGGAAAGCGAGCTGCTTTTCAGTATGACGAGCGCAAAATCTGTTGGT 1069
Oy 541 ATGCATCTGCTGCTTTCACGACGCTGCAAGAGATGCAAGCAAAATCTGCTGCA 600
Db 1070 ATGCATCTGCTGCTTTCACGACGCTGCAAGAGATGCAAGCAAAATCTGCTGCA 1129
Oy 601 GAACCGGTAATGCAAGAGATCATCAACCAATTCGCGCGCTGAATGGCTGACTTCC 660
Db 1130 GAACCGGTAATGCAAGAGATCATCAACCAATTCGCGCGCTGAATGGCTGACTTCC 1189
Oy 661 ACCAAATTCCTCATCAACCGGAGCGGCTGTTTCTGTTATCGGTGGCCAAATGGTGA 720
Db 1190 ACCAAATTCCTCATCAACCGGAGCGGCTGTTTCTGTTATCGGTGGCCAAATGGTGA 1249
Oy 721 GGTCTGACTGGTGTAAATTTATGCTGATACCTAAGCGGCAATGGCGGCTCAGCGTGG 780
Db 1250 GGTCTGACTGGTGTAAATTTATGCTGATACCTAAGCGGCAATGGCGGCTCAGCGTGG 1309
Oy 781 GGTGCTATCTCTGTGTAAGATCCATCAAAAGTGAGACCGCTTCCGAGCGCTACGACGAG 840
Db 1310 GGTGCTATCTCTGTGTAAGATCCATCAAAAGTGAGACCGCTTCCGAGCGCTACGACGAG 1369
Oy 841 TATGTCGGGAAACATCTGTTGCTGCTGCGCTGGCCATGCTTTGTAATTCAGGTTTCC 900
Db 1370 TATGTCGGGAAACATCTGTTGCTGCTGCGCTGGCCATGCTTTGTAATTCAGGTTTCC 1429
Oy 901 TACCAATCGGCGTGGCTGGAACCGACCTCCATCATGATGTAAGAAATTCGGAAGAGAAA 960
Db 1430 TACCAATCGGCGTGGCTGGAACCGACCTCCATCATGATGTAAGAAATTCGGAAGAGAAA 1489
Oy 961 GTGCTTCTGAACAAGTACCGCTGCTGTGATGAGTTCTTGACACTGGCCCATACGGT 1020
Db 1490 GTGCTTCTGAACAAGTACCGCTGCTGTGATGAGTTCTTGACACTGGCCCATACGGT 1549
Oy 1021 CTGATTTAGATGCTGATGCTGCTGACCCGATCTACAAAGAACCGGACGATACGGTAC 1080
Db 1550 CTGATTTAGATGCTGATGCTGCTGACCCGATCTACAAAGAACCGGACGATACGGTAC 1609
Oy 1081 TTTGGTGTGAACATTTCCCGTGGGAAAAAACGACAAAGGCGAGCTGCTGCCGATGCT 1140
Db 1610 TTTGGTGTGAACATTTCCCGTGGGAAAAAACGACAAAGGCGAGCTGCTGCCGATGCT 1669
Oy 1141 GCCGCTGTGAAGTAA 1155
Db 1670 GCCGCTGTGAAGTAA 1684

```

```

RESULT 4
LOCUS AX018972 4975 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 13 from Patent WO942591.
ACCESSION AX018972
VERSION AX018972.1 GI:10043067
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 4975)
AUTHORS Schroeder, H.
TITLE Method for producing bioclin
JOURNAL Patent: WO 942591-A 13 26-AUG-1999;
SCHROEDER HARTWIG (DE); BASF AG (DE)

```

```

FEATURES
    source
        Location/Qualifiers
            1..4975
                /organism="Escherichia coli"
                /db_xref="taxon:562"
                530..1684
                    CDS
                        /note="unnamed protein product"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="CAC07720.1"
                        /db_xref="GI:10043068"
                        /translation="MAKHLFTSESVSEGHPRIDQISDAVLDAILEDPPKARVACET
                        YKTKGVLVGEITTSAMVDIEITNTVREIYGVSDHGFDPANSCAVLSAIGKSPD
                        INGVDRADLPDQAGDQIMEGVATNENVDLPAPITAYFLVROAQRVNRKGLTM
                        LRPDAKSQVTPDYDQKTYGIDAVULSTHSESLDQKSLQEVAMEI IKPILPAKRLT
                        SAKREFINPTGRFVIOGPMGDCGLGTRKIIVDTVGGMAHGGAGSGKPSVDRSA
                        VARYAVAKNIIVAGLADRCIEIYSAIIGAEPSTINVENTETKEPSEQLTLVREFF
                        DLRPGLIOMLIDLHP IYKETAYGHFGREHPFKETDKAOLLRDVAAGLK"
                        1782..2987
                            /note="unnamed protein product"
                            /codon_start=1
                            /transl_table=11
                            /protein_id="CAC07721.1"
                            /db_xref="GI:10043069"
                            /translation="MNVFNPADPRAOPALODAGVYLDSATALKRPAEVAVENTOQFYS
                            LSAQVHRSOPAEAPORLTARYEAAREKVAQOLINAPDQKTIIVTRGTTESINVAOCYA
                            RRLDPGDEITVSAEIHANLYPMLHVAOQTKAKYKLELNORLPDVLPELITPR
                            SRLAGQMSNVTVGCTVDLARAITTFANSHGVMVMQDAGAVHFADVOULDIPTAF
                            SGHKLPGLEIVLYGXSELLAMSPMLGKVMHVSFGFTTOSAPKPLEGTPNV
                            AGVITGLSALEMLADYDINOASWSRSLTYTLAEDALAKRPFRRFCQDSLIADFPA
                            GVHSDMTVLAEYGIYLRAGQICAPFLAELGVGTTLRASPAPYNTKSDVALLVNAV
                            DRAELLYVO"
    CDS
        BASF COUNT 1198 a 1350 c 1329 g 1098 t
        ORIGIN
            Query Match 100.0%; Score 1155; DB 6; Length 4975;
            Best Local Similarity 100.0%; Pred. No. 1.2e-266;
            Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y	541	ATCCATGCTGTCGTCGCTTCCATCAGCAGCAGTGTGAAGAGATCGACACAAATTCGCTGAA	600
Db	1070	ATCGATGCTGTCGTCGCTTCCATCAGCAGCAGTGTGAAGAGATCGACACAAATTCGCTGAA	1129
Y	601	GAAGCCGTAATGGAAGATCATCAAGCAATTCGCCCCGTAATGAGTGAATTCGCTGCC	660
Db	1130	GAAGCCGTAATGGAAGATCATCAAGCAATTCGCCCCGTAATGAGTGAATTCGCTGCC	1189
Y	661	ACCAAATTCCTTCATCAACCCGACCGGTCGTTTCGTTATTCGGTGGCCCAATGGGTACTGC	720
Db	1190	ACCAAATTCCTTCATCAACCCGACCGGTCGTTTCGTTATTCGGTGGCCCAATGGGTACTGC	1249
Y	721	GGTGTGACTGGTCGTAATATTCGTTGATACCTACGGCGGATGGCCGTCACGCTGAC	780
Db	1250	GGTGTGACTGGTCGTAATATTCGTTGATACCTACGGCGGATGGCCGTCACGCTGAC	1309
Y	781	GGTCGATTCCTCTGGTAAAGATCATCAAAAGTGACGCTCCGACGCTACGACGACGT	840
Db	1310	GGTCGATTCCTCTGGTAAAGATCATCAAAAGTGACGCTCCGACGCTACGACGACGT	1369
Y	841	TATGTCCGGAANAACATCGTTCGTCGTCGCTGGCCGATCGTTGTGAATTCAGGTTCC	900
Db	1370	TATGTCCGGAANAACATCGTTCGTCGTCGCTGGCCGATCGTTGTGAATTCAGGTTCC	1429
Y	901	TACGCAATCGGCGTGGCGTGAACGACGCTCCATCATGTAGAAACTTTCGTAAGAGAA	960
Db	1430	TACGCAATCGGCGTGGCGTGAACGACGCTCCATCATGTAGAAACTTTCGTAAGAGAA	1489
Y	961	GTGCTTCTGGAACAACGACCTGCTGTAGCTGAGTTCCTTGACCTCGCCCATACGCT	1020
Db	1490	GTGCTTCTGGAACAACGACCTGCTGTAGCTGAGTTCCTTGACCTCGCCCATACGCT	1549
Y	1021	CTGATTCAGATCGTCGATCTGCTGCACCCGATCTTCAAAAGAAACCGCAGCATACGCTAC	1080
Db	1550	CTGATTCAGATCGTCGATCTGCTGCACCCGATCTTCAAAAGAAACCGCAGCATACGCTAC	1609
Y	1081	TTTGGTCGTGAACAATTTCCCTGGGAAAAAACCGACAAGCGCAGCTGCGCGATGCT	1140
Db	1610	TTTGGTCGTGAACAATTTCCCTGGGAAAAAACCGACAAGCGCAGCTGCGCGATGCT	1669
Y	1141	GGCGGTCTGAAGTAA 1155	
Db	1670	GGCGGTCTGAAGTAA 1684	
RESULT 5			
AE000377			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
MEDLINE			

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
608-263-7459 3 (bases 1 to 12354) Blattner F.R. Direct Submission Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	4 (bases 1 to 12354) Plunkett G. III Direct Submission Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA On Sep 9, 1997 this sequence version replaced gi:1789304. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCRR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site NOS., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). **The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms in Entrez as gene labels. This should allow them to be searched for in Entrez as gene names.	
FEATURES source	Location/Qualifiers 1..12354 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:33333" complement(39..67) /note="factor sigma70; predicted +1 start at 3079733"	
promoter	105..989 /gene="yggG" /note="b2936"	
gene	103..989 /gene="yggG" /function="orf; unknown" /note="0294; 100 pct identical to YGGC_ECOLI SM: P25894" /codon_start=1 /trans_table=1 /product="orf, hypothetical protein" /protein_id="AAC75973.1" /db_xref="GI:1789105" /translation="MLACCAATSPVRCAPPTTEREFTCKRRIPLILTCGKNNKK IRALLVSVATVLTGCGNDSNGLSSGAEPAVSLDQVLTSDQCGEDSKAA TIRANSEYAKRLTTINALGNINICQPVYKYMAMDVNAFMANCIRYSCILGM KTLNVEFAVIGHEGVALGKVGKQGVYALCTNAVRVAAASAGCIVSLSOLGNGK EKLVNSQFSGQEQEAEDVYSYDLRLRGISPAGLATFEXKLAEKROSSMHDHPRA SAERAQHIRDRMSADCIK"	
CDS	1008..1154 /note="REP (repetitive extragenic palindromic) element; contains 3 REP sequences" complement(1195..2115) /gene="spB" /note="b2937"	
repeat_region		
gene		

CDS
complement(1195..2115)
/gene="speB"
/EC_number="3.5.3.11"
/function="enzyme: Central Intermediary metabolism: Polyamine biosynthesis"
/note="f306: 100 pct identical to speB_ECOLI SW: P16936; CG Site No. 160"
/codon_start=1
/transl_table=11
/product="agmatinase"
/protein_id="AAC75974.1"
/db_xref="GI:1789306"
/translation="MSTLGHQDNLVSNAGFLRLPMNPQYDSDAVYITGVPEM ATSGRACGRGPAIRQVSNLAMEHNPMPNMDRELRNVCCDGLVAFCDAREMS EKLDHAEKLLAGKRLSTGDFVTLPLRLHAKHFGMALVHEDARTDTANGCE FDHGTMYTAPKEGLIDPNHSVQIGIRTEFDKNGKFTVDACQNDNSVDVIAQVKE IVDGMYTFLFDIDCLDPAFAPGSGTPIVIGLSDRAIKLVRLKLDKNTIVGMVVEVA PAYDSEITLAAATLALMLYIOAAKGE"
complement(2253..4229)
/gene="speA"
/note="b2938"
complement(2253..4229)
/gene="speA"
/EC_number="4.1.1.19"
/function="enzyme: Central intermediary metabolism: Polyamine biosynthesis"
/note="f658: 99 pct identical to speA_ECOLI SW: P21170; CG Site No. 161"
/codon_start=1
/transl_table=11
/product="biosynthetic arginine decarboxylase"
/protein_id="AAC75975.1"
/db_xref="GI:1789307"
/translation="MSDDMSGLPSSAGEBVLRSQMOEAMSSQESKMLRTYNIAM GNNVYDNLGHSVCPDVPPEARVDLQVTRQOGRPLALCFEPOILQRLHS INAAFKRARESYNGDPLVYPKVNQHRVIESLHSGEPLEAGSKAELMAYLA HAGTRSVIANGYKDKREYIRLALIGEMGHKYLVEKSEIAYIDEAERLVNPR LGVBARLASQSGMOSSGGEKSKFLAATQVLOLVETLEAGRLDLSLHRLHSGQ MANIRDIATGVRASARVYELHKLVGNICQFQVGGCIGVDEGRSQCSDSVAGNE YANNITWATGDACEENCLPHPTVITTESGRAVTAHHVTVLVNITIGVENETVTPAPE DAPRALQSMETQEMHEPGRTRSLREMLDSQMDLHIIIVSSGIFSLQERMAQO LVLSMCEHVOKOLDPONRAHPIIDLOEIMADGMVNFSLFQSPMPANGIDOLFVPL PLEGLDVPERRALVDITCSDGAIIDHYIDMGKATITMPEVDMPNMLFFVNG AYOELIGMHNILFGDPEADVAVFVPGSGVLELSDSDTVADMLOYQDLPRKTLTQF RDQVKRTDLDAELQQCFLEFEAGLYVTYLTDEB"
complement(2271..2297)
/gene="speA"
/note="factor Sigma70: promoter speB: documented +1 at 3081965"
2759..2791
/note="REP (repetitive extragenic palindromic) element: contains 1 REP sequence"
complement(4238..4384)
/gene="yggB"
/note="b2939"
complement(4238..4384)
/gene="yggB"
/function="orf: Unknown"
/note="f48: 100 pct identical to yggB_ECOLI SW: P46877"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75976.1"
/db_xref="GI:1789308"
/translation="MLNNVMKKRPVQALEHROHSILENPCAAYGLLSQFOAAIVNCFLL NKII"
complement(4272..4300)
/gene="yggB"
/note="factor Sigma70: predicted +1 start at 3083966"
4339..4368
/note="factor Sigma70: predicted +1 start at 3084076"
complement(4469..4496)
/note="factor Sigma70: predicted +1 start at 3084163"

gene
4505..4720
/gene="yggC"
/note="b2940"
4505..4720
/gene="yggC"
/function="orf: Unknown"
/note="o71: 100 pct identical to yggC_ECOLI SW: P46878"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75977.1"
/db_xref="GI:1789309"
/translation="MGITSAQMSQSDAEGGERIFTRTVROVKQOTTVHYFVSPRPV KTRPQKTLISITPLEVATRKRVLLI"
complement(4717..4968)
/gene="yggD"
/note="b2941"
complement(4717..4968)
/gene="yggD"
/function="orf: Unknown"
/note="f83: 100 pct identical to yggD_ECOLI SW: P46879"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75978.1"
/db_xref="GI:1789310"
/translation="WVADPTTLQVNTGSLSYRKYGINIMWAILQGFLLFPLFFE SCULIKTWLEIFPDNAGILRLIYLLDQSAIVGKTRRAA"
4849..4878
promoter
Query Match 100.0%; Score 1155; DB 1; Length 12354;
Best Local Similarity 100.0%; Pred. No. 1; 3e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
gene
5024 ATGCAAAACACCTTTTACGTCGAGTCCGTCCTGAAGGCAATCTGACAAATTGCT 60
1 ATGCAAAACACCTTTTACGTCGAGTCCGTCCTGAAGGCAATCTGACAAATTGCT 60
5084 GACCAAAATTTCTGATGCGCTTTTACGTCGAGTCCGTCCTGAAGGCAATCTGACAAATTGCT 120
61 GACCAAAATTTCTGATGCGCTTTTACGTCGAGTCCGTCCTGAAGGCAATCTGACAAATTGCT 120
5084 GACCAAAATTTCTGATGCGCTTTTACGTCGAGTCCGTCCTGAAGGCAATCTGACAAATTGCT 120
121 GCTTGGCAAAACCTCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 180
GCTTGGCAAAACCTCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 180
5144 GCTTGGCAAAACCTCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 240
181 GCTTGGCAAAACCTCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 240
5204 GCTTGGCAAAACCTCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 240
241 TCGGCAATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 300
TCGGCAATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 300
5264 TCGGCAATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 360
301 CCGGCAATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 360
5324 CCGGCAATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 420
361 GGTGATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 420
5384 GGTGATGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 480
421 TATGCAACGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 540
5444 TATGCAACGCTGTAAGGCAATCTGACAAATTGCTGACAAATTGCT 540
481 CTGGCCCGGCAAGGCAATCTGTAAGGCAATCTGACAAATTGCT 540
5504 CTGGCCCGGCAAGGCAATCTGTAAGGCAATCTGACAAATTGCT 540
541 ATGATGCTGTAAGGCAATCTGTAAGGCAATCTGACAAATTGCT 600
5564 ATGATGCTGTAAGGCAATCTGTAAGGCAATCTGACAAATTGCT 600

```

OY 601 GAAGCGTAATGGAAGAGATCATCAAGCAATTTGCGCGGTGAATGCTGACTTGTGCC 660
DB 5624 GAAGCGTAATGGAAGAGATCATCAAGCAATTTGCGCGGTGAATGCTGACTTGTGCC 5683
OY 661 ACCAAATTTCTTCATCAACCCGACCGGTGCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
DB 5684 ACCAAATTTCTTCATCAACCCGACCGGTGCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743
OY 721 GGTCTGACGTGCTGTAATAATTTATCGTTGATACCTAGCGGGGCATGGCGGTGCTGACGTGCG 780
DB 5744 GGTCTGACGTGCTGTAATAATTTATCGTTGATACCTAGCGGGGCATGGCGGTGCTGACGTGCG 5803
OY 781 GGCGATTCTCTGTAATAATCATCAAGAGTGGACGTTCCGACCGCTACCGAGCAGCT 840
DB 5804 GGTGCAATTTCTGTAATAATCATCAAGAGTGGACGTTCCGACCGCTACCGAGCAGCT 5863
OY 841 TATGTGCGCAAAAACATCTGCTGCTGCGCGCGCATGCTTGTGAATAATTCAGGTTTCC 900
DB 5864 TATGTGCGCAAAAACATCTGCTGCTGCGCGCGCATGCTTGTGAATAATTCAGGTTTCC 5923
OY 901 TACGCAATGGCGGTGCTGCTGCTGCAACCGACCTTCATCATGTGTGAATACTTGGGTACTGAGAA 960
DB 5924 TACGCAATGGCGGTGCTGCTGCTGCAACCGACCTTCATCATGTGTGAATACTTGGGTACTGAGAA 5983
OY 961 GTGCGCTTCTGAACAACATGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 5984 GTGCGCTTCTGAACAACATGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6043
OY 1021 CTGATTTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 6044 CTGATTTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6103
OY 1081 TTTTGTCTGCAACATTTTCCCGTGGGAAAAAACCGACGACGCTGCTGCGCGCTGCT 1140
DB 6104 TTTTGTCTGCAACATTTTCCCGTGGGAAAAAACCGACGACGCTGCTGCGCGCTGCT 6163
OY 1141 GCCGCTCTGAGTAA 1155
DB 6164 GCCGCTCTGAGTAA 6178

RESULT 6
AX370209 12354 bp DNA linear PAT 16-FEB-2002
DEFINITION Sequence 26 from Patent WO0170776.
ACCESSION AX370209
VERSION AX370209.1 GI:18857386
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1
AUTHORS Levy, S.B., Barbosa, T.M. and Aleksun, M.N.
TITLE Nmr compositions and their methods of use
JOURNAL Patent: WO 0170776-A 26-27-SEP-2001;
TRUSTEES OF TUFTS COLLEGE (US)
FEATURES
SOURCE
1. 12354
/organism="Escherichia coli"
/db_xref="taxon:562"
10196..11146
/note="unnamed protein product"
/codon_start=1
/translation="1"
/protein_id="CAD23853.1"
/db_xref="GI:18857387"
/translation="MIKGIYVDPININIKDSSFAMLEAORRCYELHYKENDLY
LNGEAAHRTLVNKNVEFVSGODPLADLVILMKRDPFDEFTIYATYL
BRAEKGILIVKPSLRDCKNEKLTPTPELTITRNKAQDKATWEKSIDILK
PLDGMGASIFVKEGDPNLGVIETLEHGRYCAONVYPAIKDCKRVLVYGEF
VPYCLARIPOGGETRGNLAAGRGEPRLTESDMKIAIOIGTLKGLIFVGLDITG

```

```

BASE COUNT 3031 a 3249 c 3190 g 2884 t
ORIGIN
Query Match 100.0%: Score 1155; DB 6; Length 12354;
Best Local Similarity 100.0%: Prod. No. 1.38-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGCAAAACACCTTTTACGTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
DB 5024 ATGGCAAAACACCTTTTACGTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5083
OY 61 GACCAATTTCTGATGCGCTTTTACGACGATCTTGAAGAGATTCGGAACACGCGTT 120
DB 5084 GACCAATTTCTGATGCGCTTTTACGACGATCTTGAAGAGATTCGGAACACGCGTT 5143
OY 121 GCTTGGCAAAACCTACGTAATAAACCGGCAATGTTTATGTTGGCGGCAAAATTCACCCAGC 180
DB 5144 GCTTGGCAAAACCTACGTAATAAACCGGCAATGTTTATGTTGGCGGCAAAATTCACCCAGC 5203
OY 181 GCTTGGCAAAACCTACGTAATAAACCGGCAATGTTTATGTTGGCGGCAAAATTCACCCAGC 240
DB 5204 GCTTGGCAAAACCTACGTAATAAACCGGCAATGTTTATGTTGGCGGCAAAATTCACCCAGC 5263
OY 241 TCCGACATGGCGCTTGGACGCTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 5264 TCCGACATGGCGCTTGGACGCTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5323
OY 301 CCTGACATCAACGAGGCGCTTGAACCGTGCATCCGCTGGAACAGGCGCGGTGACAG 360
DB 5324 CCTGACATCAACGAGGCGCTTGAACCGTGCATCCGCTGGAACAGGCGCGGTGACAG 5383
OY 361 GGTCTGATGTTTGGCTACCGCACTAATGAACCGAGCTGCTGATGCGACACTATACAC 420
DB 5384 GGTCTGATGTTTGGCTACCGCACTAATGAACCGAGCTGCTGATGCGACACTATACAC 5443
OY 421 TATGACACGCTGCTGTAAGCGTACGCTGTAAGTGCCTGTAAGAACGCGACTGCGCGT 480
DB 5444 TATGACACGCTGCTGTAAGCGTACGCTGTAAGTGCCTGTAAGAACGCGACTGCGCGT 5503
OY 481 CTGCGCGCGGACGCAAAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 5504 CTGCGCGCGGACGCAAAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5563
OY 541 ATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 5564 ATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5623
OY 601 GAAGCGTAATGGAAGAGATCATCAAGCAATTTGCGCGGTGAATGCTGACTTGTGCC 660
DB 5624 GAAGCGTAATGGAAGAGATCATCAAGCAATTTGCGCGGTGAATGCTGACTTGTGCC 5683
OY 661 ACCAAATTTCTTCATCAACCCGACCGGTGCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
DB 5684 ACCAAATTTCTTCATCAACCCGACCGGTGCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743
OY 721 GGTCTGACGTGCTGTAATAATTTATCGTTGATACCTAGCGGGGCATGGCGGTGCTGACGTGCG 780
DB 5744 GGTCTGACGTGCTGTAATAATTTATCGTTGATACCTAGCGGGGCATGGCGGTGCTGACGTGCG 5803
OY 781 GGCGATTCTCTGTAATAATCATCAAGAGTGGACGTTCCGACCGCTACCGAGCAGCT 840
DB 5804 GGTGCAATTTCTGTAATAATCATCAAGAGTGGACGTTCCGACCGCTACCGAGCAGCT 5863
OY 841 TATGTGCGCAAAAACATCTGCTGCTGCGCGCGCATGCTTGTGAATAATTCAGGTTTCC 900
DB 5864 TATGTGCGCAAAAACATCTGCTGCTGCGCGCGCATGCTTGTGAATAATTCAGGTTTCC 5923
OY 901 TACGCAATGGCGGTGCTGCTGCTGCAACCGACCTTCATCATGTGTGAATACTTGGGTACTGAGAA 960
DB 5924 TACGCAATGGCGGTGCTGCTGCTGCAACCGACCTTCATCATGTGTGAATACTTGGGTACTGAGAA 5983
OY 961 GTGCGCTTCTGAACAACATGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

```

|||||
Db 5984 GTGCTTCTGACACATGACCTGCTGCTAGCTAGTTCTTCCAGCTGCCGCCCTATACGT 6043
OY 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCTACAAAGAACCCGAGCATACGCTCAC 1080
|||||
Db 6044 CTGATTCAGATGCTGATCTGCTGACCGCATCTACAAAGAACCCGAGCATACGCTCAC 6103
OY 1081 TTGCGTGTGAACATTTCCGCTGGGAAAAACGACAAACCGCAGCTGCTGCGGATGCT 1140
|||||
Db 6104 TTGCGTGTGAACATTTCCGCTGGGAAAAACGACAAACCGCAGCTGCTGCGGATGCT 6163
OY 1141 GCCGCTCTGAAGTAA 1155
|||||
Db 6164 GCCGCTCTGAAGTAA 6178
RESULT 7
AX370245 12354 bp DNA linear PAT 16-FEB-2002
LOCUS Sequence 62 from Patent WO0170776.
DEFINITION AX370245
VERSION AX370245.1 GI:18857412
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1
AUTHORS Levy, S.B., Barbosa, T.M. and Aleksun, M.N.
TITLE Nmr compositions and their methods of use
JOURNAL Patent: WO 0170776-A 62 27-SEP-2001;
TRUSTEES OF TUFTS COLLEGE (US)
FEATURES
Source Location/Qualifiers
1..12354
/organism="Escherichia coli"
/db_xref="taxon:562"
1183..11818
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD23861.1"
/db_xref="GI:18857413"
/translation="MTAPFPPRLGAVAFLENOETPLTNLHFLIAMPALODPIFR
RSVYICENHTNGAMGIIYKPLENLIKIGLIEKLTPEPRPESRLDKPVYMGPL
AEDRGFIHTPPSNFASIRISDNTVMTTSRDVLELTGDKDPSDVLVAGVSWKMG
OLEDEILDNAWLTAPADNLILFTPTPIADWREAKLIGVDILTMPGVAGHA"
BASE COUNT 3031 a 3249 c 3190 g 2884 t
ORIGIN
Query Match 100.0%; Score 1155; DB 6; Length 12354;
Best Local Similarity 100.0%; Pred. No. 1.3e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 301 CTTGACATCAACAGGCGCTTGACCGT3CCGATCCGCTGGAACAGGCGCGGCTGACAG 360
Db 5324 CTTGACATCAACAGGCGCTTGACCGT3CCGATCCGCTGGAACAGGCGCGGCTGACAG 5383
OY 361 GGTCTGATGTTTGGCTACGCACTAATGAACGAGTGTGATGCGACACCTATCAC 420
Db 5384 GGTCTGATGTTTGGCTACGCACTAATGAACGAGTGTGATGCGACACCTATCAC 5443
OY 421 TATGACACCGTGTGATGACGCTCAAGCTGAAGTGCCTAAACGCGACCTTGCCGTG 480
Db 5444 TATGACACCGTGTGATGACGCTCAAGCTGAAGTGCCTAAACGCGACCTTGCCGTG 5503
OY 481 CTGCGCCCGGACGGAACAGCGTGAAGTGCCTAAACGCGACCTTGCCGTG 540
Db 5504 CTGCGCCCGGACGGAACAGCGTGAAGTGCCTAAACGCGACCTTGCCGTG 5563
OY 541 ATGATGCTGTGCTGCTTTCACCTAGCACTGTAAGAGATGACCAAGAAATGCTGCA 600
Db 5564 ATGATGCTGTGCTGCTTTCACCTAGCACTGTAAGAGATGACCAAGAAATGCTGCA 5623
OY 601 GAAGCGGTAATGAGAGATCATCAAGCAATTCGCCGCTGAATGCTGACTTCTGC 660
Db 5624 GAAGCGGTAATGAGAGATCATCAAGCAATTCGCCGCTGAATGCTGACTTCTGC 5683
OY 661 ACCAAATTCCTCAACCCGACCGGCTGCTTATGCTGCGCCCAATGGGTGACTGC 720
Db 5684 ACCAAATTCCTCAACCCGACCGGCTGCTTATGCTGCGCCCAATGGGTGACTGC 5743
OY 721 GGTCTGCTGCTGCTGTAATATATGTTGATACCTAGCGGCGCATGCGCGCTCACGCTGC 780
Db 5744 GGTCTGCTGCTGCTGTAATATATGTTGATACCTAGCGGCGCATGCGCGCTCACGCTGC 5803
OY 781 GGTGATTTCTGCTGTAAGATCCATCAAAAGTGACCGTTGCGACCTACGACGACGT 840
Db 5804 GGTGATTTCTGCTGTAAGATCCATCAAAAGTGACCGTTGCGACCTACGACGACGT 5863
OY 841 TATGCGCGAAACATCGTGTGCTGCGCGCTGCGGATCGTTGTAATTCAGTTTCC 900
Db 5864 TATGCGCGAAACATCGTGTGCTGCGCGCTGCGGATCGTTGTAATTCAGTTTCC 5923
OY 901 TACGCAATCGCGCTGCTGACCGACCTCCATCATGTAAGAACTTTCGCTACTGAGAA 960
Db 5924 TACGCAATCGCGCTGCTGACCGACCTCCATCATGTAAGAACTTTCGCTACTGAGAA 5983
OY 961 GTGCGCTTGAACACTGACCTGCTGTAAGTGTGCTTTCGACCTGCGCCCATACGGT 1020
Db 5984 GTGCGCTTGAACACTGACCTGCTGTAAGTGTGCTTTCGACCTGCGCCCATACGGT 6043
OY 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCTACAAAGAACCCGAGCATACGCTCAC 1080
Db 6044 CTGATTCAGATGCTGATCTGCTGACCGCATCTACAAAGAACCCGAGCATACGCTCAC 6103
OY 1081 TTGCGTGTGAACATTTCCGCTGGGAAAAACGACAAACCGCAGCTGCTGCGGATGCT 1140
Db 6104 TTGCGTGTGAACATTTCCGCTGGGAAAAACGACAAACCGCAGCTGCTGCGGATGCT 6163
OY 1141 GCCGCTCTGAAGTAA 1155
Db 6164 GCCGCTCTGAAGTAA 6178
RESULT 8
AX370258 12354 bp DNA linear PAT 16-FEB-2002
LOCUS Sequence 75 from Patent WO0170776.
DEFINITION AX370258
ACCESSION AX370258
VERSION AX370258.1 GI:18857424
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1

AUTHORS Levy, S.B., Barbosa, T.M. and Alekshun, M.N.
 TITLE Nmr compositions and their methods of use
 JOURNAL Patent: WO 0170776-A 75 27-SEP-2001;
 TRUSTEES OF TUFTS COLLEGE (US)

FEATURES
 source Location/Qualifiers
 1..12354
 /organism="Escherichia coli"

CDS
 9425..10183
 /note="unnamed protein product"

/db_xref="taxon:562"

/codon_start=1

/transl_table=1

/protein_id="CAD3867.1"

/db_xref="GI:18857425"

/translation="MGRRLGLTMRIRIYHPEPLTSHSLCEADANHIGVLRMG
 PGOLQLFDGNSVYQDAITRSKSKSVKVEKLEIDIDREPLIHLCQVRSREKHE
 FTTOKSIELVSLITPLSEKGVKLDSEKLVNKLQOMKJAIACEGKRNRPETR
 PAMLEAMCAEDGDLKLNLPKASNSINLPLEVERRLIGEGGLSADDEIAMTAR
 YQFDILGPVLRTEETLTALTAITAIQVRFGLG"

BASE COUNT 3031 a 3249 c 3190 g 2884 t

ORIGIN

Query Match 100.0%; Score 1155; DB 6; Length 12354;

Best Local Similarity 100.0%; Pred. No. 1,3e-266;

Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGAAACACCTTTTACGTCGAGTCCTCTGAAAGGCATCCGACAAATTCCT 60
 DB 5024 ATGGGAAACACCTTTTACGTCGAGTCCTCTGAAAGGCATCCGACAAATTCCT 5083
 OY 61 GACCAAAATTCGATCCGCTTTTACGATCCTGGAAGATCCGAAAGCAGCGCT 120
 DB 5084 GACCAAAATTCGATCCGCTTTTACGATCCTGGAAGATCCGAAAGCAGCGCT 5143
 OY 121 GCTTCGAAACCTACGTAAACCGGCATGTTTATGTTGGCGGGAATTCACACAC 180
 DB 5144 GCTTCGAAACCTACGTAAACCGGCATGTTTATGTTGGCGGGAATTCACACAC 5203
 OY 181 GCGTGGTAGACATGGAAGAGATCAACCGGTACACCGTTGCGGAATTCGATGTCAT 240
 DB 5204 GCGTGGTAGACATGGAAGAGATCAACCGGTACACCGTTGCGGAATTCGATGTCAT 5263
 OY 241 TCCGACATGCGCTTACGCTAATCTGTCGCTTCTGAGCGCTATCGGCAACAGCT 300
 DB 5264 TCCGACATGCGCTTACGCTAATCTGTCGCTTCTGAGCGCTATCGGCAACAGCT 5323
 OY 301 CCTGACATCAACGAGCGCTTACCGTCCGATCCGCTGGAACAGGCGCGGTACAC 360
 DB 5324 CCTGACATCAACGAGCGCTTACCGTCCGATCCGCTGGAACAGGCGCGGTACAC 5383
 OY 361 GGTGATGATGCTGCTGACGACACTAATGAACGAGCGTCTATGCGACACCTATCAG 420
 DB 5384 GGTGATGATGCTGCTGACGACACTAATGAACGAGCGTCTATGCGACACCTATCAG 5443
 OY 421 TATGACACCGTCTGCTAGCAGCTGAGCTGGAAGTCCGTAACAGGCACTCTGCCGTG 480
 DB 5444 TATGACACCGTCTGCTAGCAGCTGAGCTGGAAGTCCGTAACAGGCACTCTGCCGTG 5503
 OY 481 CTGCGCGCGGACGCAAAAGCAGTCTTTCAGTATGACGAGCGCAAAATCGTTGCT 540
 DB 5504 CTGCGCGCGGACGCAAAAGCAGTCTTTCAGTATGACGAGCGCAAAATCGTTGCT 5563
 OY 541 ATCGATGCTGCTGCTTTCACACTGACACTCTGAAGAGATGACACGAAATCGCTGCA 600
 DB 5564 ATCGATGCTGCTGCTTTCACACTGACACTCTGAAGAGATGACACGAAATCGCTGCA 5623
 OY 601 GAAGCGTATGAGAGATCATCAAGCAATTCGCCGCTGAATGCTGATCTTGGC 660
 DB 5624 GAAGCGTATGAGAGATCATCAAGCAATTCGCCGCTGAATGCTGATCTTGGC 5683
 OY 661 ACCAAATTCATCAACCGAGCGCTTTCGTTATGCTGCGCCCAATGGCTGACTGC 720
 DB 5684 ACCAAATTCATCAACCGAGCGCTTTCGTTATGCTGCGCCCAATGGCTGACTGC 5743

OY 721 GGTCTGAGTGGTGTGAATAATATGCTTGATACCTACGGCGCATGGCGCTACGGTGC 780
 DB 5744 GGTCTGAGTGGTGTGAATAATATGCTTGATACCTACGGCGCATGGCGCTACGGTGC 5803
 OY 781 GGTCTGATCTCTGTGAAGATTCATCAAAAGTGAACCTTCCGACGCTACGACGACT 840
 DB 5804 GGTCTGATCTCTGTGAAGATTCATCAAAAGTGAACCTTCCGACGCTACGACGACT 5863
 OY 841 TATGTCGGGAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 DB 5864 TATGTCGGGAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5923
 OY 901 TACGCAATCGCGTGGCTGAACCGACCTCCATCATGATGATGAAGTTCGCTACTAGAAA 960
 DB 5924 TACGCAATCGCGTGGCTGAACCGACCTCCATCATGATGATGAAGTTCGCTACTAGAAA 5983
 OY 961 GTGCTTGTGAACACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 5984 GTGCTTGTGAACACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6043
 OY 1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 6044 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6103
 OY 1081 TTTGCTGTGAACATTTCCGCTGGGAAACCGCAAGGCGCAGCTCTGCGGATGCT 1140
 DB 6104 TTTGCTGTGAACATTTCCGCTGGGAAACCGCAAGGCGCAGCTCTGCGGATGCT 6163
 OY 1141 GCGGCTGTGAAGTAA 1155
 DB 6164 GCGGCTGTGAAGTAA 6178

RESULT 9

ECU28377
 LOCUS ECU28377 141744 bp DNA linear BCT 01-JUL-1995
 DEFINITION Escherichia coli K-12 genome; approximately 65 to 68 minutes.
 ACCESSION U28377
 VERSION U28377.1 GI:882431
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 87107)

AUTHORS Plunkett, G.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. This entry should be considered somewhat provisional; it will be updated and merged with others at a later date.

FEATURES
 source Location/Qualifiers
 1..141744
 /organism="Escherichia coli"

/strain="K-12"

/sub_strain="MG1655"

/db_xref="taxon:562"

/map="approximately 65 to 68 minutes"

/note="This sequence comprises the following lambda clones: DB656(EC18-164), DB658(EC17-126), DB660(EC30M49), DB663(EC19-62), DB665(EC27-6896), DB667(EC27-357), DB670(EC30-45), DB674(EC27-1001), DB676(EC27-1160), DB678(EC30K505-2), DB680(EC30K506-2), and DB682(EC17-140); M13Janus vector was used for subcloning"

misc_feature complement(1..535)
 /note="similar to GenBank Accession Number X73958"

misc_feature	CDS	(ECCGVOP) complement(1. .292) /note="ORF_f97p" /codon_start=1 /transl_table=11 /protein_id="AA69070.1" /db_xref="GI:882432" /translation="MAIALVTVGGSGRIGRATALLAOGGYTVAVNYOONLNAOEVM LITQAGKAFVLQADISDENOVVAMFTIIDHDEPLALVNNAGILFTQCTIVE" complement(70. .324) /note="similar to GenBank Accession Number D21144 (ECOPYU18)"
		complement(318. .3634) /note="similar to GenBank Accession Number L20872 (ECOCVVP). bases 3318 to 3438 differ from the sequence presented here, as well as from X73958 and D21144" complement(559. .3432) /gene="gcvp" complement(559. .3432) /note="ORF_f957" /codon_start=1 /transl_table=11 /protein_id="AA69071.1" /db_xref="GI:882433" /translation="MTOTLSOLENSGAFIERHIGPDAAOQOENLNAVQOSLNLTQO IYPKDIQIATPPQVACAPATEVAALAEIKAIASRNKRTFSYIGMGYTAQVLPVILRM LEPCMTATATPYOPEVSOGRLLEALLRFOQVTLDTGLDMASASLIDETTAAEAMAM ARKVKKLNKANEFEVASDVPQTLVVRATFEFEVIVDQKVLDHODVGYLLO QVGTGEIHDTYALISELSKRIIVSVADIADVALVLTAPKQADIVFGSAQREKSVI MGYGPFAAFKADDEYKSRMGRRIIVSKDAAGNTALMAQTEQOIRRKANSNI CPSOVLNIALSVAVHGPVGLKLRANIRHRLDTLALGLOQKGLKLRHAFYDLC VEVADKAGYLTRAEAELNLRSDILANVGTITDETTRENNWOLFNVLLDGHGIDID TLDKADHSRSIOPMLRDETLTHVPVNRHSETEMKRYHMSLEKRLALNOAMIP LSCSMKLNAAAEMLPTWPEAEHLPCPPQAEYQOQIAQLADWLKGLGIDVAC MOPNSGAOGEVAGLLAIRHYHESRNEGHRDICIIPASAGTNPASAHMAGOVVAVAC DKNGNIDLTBAKAEQAGDNLSCIVTPTSTHGYEETIREVCEVYHQFGQVLDG AMNNAOVGITSPTFGADVSHLNLKHTFCIPHGGGPGKPIGVKAHLAPFYGSHV QIECHITRGCAYSAAPGCSATIPISMYIRMMGAGLKAKASOVAILMNYITFASRTLD ATPVLITGRGVNAHECILDIRPLKEETISLIDIAKRILIDVGFNAPTNTSPACITLM VPTESKSEVELDRFDAMLAIRAEIDOVKACVWPLEDPLVNAPIHOSLEVAEMAH VSREVAVFPAGVADKXWPTVKRLDDVYGDRLNFCSCVPISEYO" complement(3551. .3940) /gene="gcvh" complement(3551. .3940) /note="ORF_f129" /codon_start=1 /transl_table=11 /protein_id="AA69072.1" /db_xref="GI:882434" /translation="MSNPAELKYSKEHEMLKREADCTYTGITTEHAOELGDMVFD LPEVGVTVSAGDCAVAESVKAASDIYAPVSGEIVAVNDALSDSPELVNSEPAGMI FKIKASDESELESLDATTAYEALLEDE" complement(3809. .5631) /note="similar to GenBank Accession Number M97263 (ECOTPRO)"
misc_feature	CDS	complement(3964. .5058) /gene="gcvt" complement(3964. .5058) /note="gcvt" /note="ORF_f364" /codon_start=1 /transl_table=11 /protein_id="AA69073.1" /db_xref="GI:882435" /translation="MAOQPLVEQHTLCGARVDFHGMMPPLHYGSOIDEHNAVRTDA GMFDVSHMTIYDLRGSRTRRELYLLANDYAKLTJSGKALYSGMLNAGSGVITDILIV YFEDFRLVNVNATREKDLSTIOTHPGEGEITTYADLPSMIAVGPNAQAATLFL NDAQRAVEGMKPPFVQAGDLFIATGTGEGEYALPNEKAADWRWLVAGVAK CGLGARDTLRLGAGMNLVYCOEMDETISPLANNGWLIAMPADRDFIGRALVGVKRP GTEKLVGTVMTKEGVLRNELPVRFTDOAGQHGEGIIITSCFSPTLGSIALAVPEGI GETAIVQIRREMPKVKYTKPVFNCKAVAV"
		complement(5370. .10317) /note="similar to GenBank Accession Number D90281 (ECOPYU)" complement(5506. .6708) /gene="v1sc" complement(5506. .6708) /gene="v1sc" /note="ORF_f400" /codon_start=1 /transl_table=11 /protein_id="AA69074.1" /db_xref="GI:882436" /translation="MQSVDAIVGGVAVGLAVACGCGSLRVAVLEORVPEPLANA PPOLRVSAINASEKLLRLGVMDILSRPASCYGVMEVMDKDSFGHISFDOSMGVS HUGHIVENSIVHYALMRKHAOSDITLLAPLELOVANGENETFLTLKDGSMILARLV IGADGNSVILNKADIPLTFMDYOHALLVATIRTEEPHDAAVARYHGEGLIAPLPS DPHICSYMSISPEAOBMOOASEDEFERANIALRDNELGCKVESAROVPLTGRVA RQFASHRLALVGDAHRTIHPLAGOSVNGPNDAAELIAELKRIHROGCDIOXYIARR YERSRHSAAALMLAGMGFRDLFGSTNPAAKLLRDLTGLADTLPGVAKPOLIRQAMGL NDLPEMLR" complement(6731. .7909) /gene="ubiH" complement(6731. .7909) /note="ubiH" /codon_start=1 /transl_table=11 /protein_id="AA69075.1" /db_xref="GI:882437" /translation="MSVIIIVGGAGATLALAIISPLSGALPVHLIETAVPESIAHPC EDGRAIALAAGCQQLARIGWQSLADCATATITVHVSDRGHACFTTAAEDYQALAL GGVVALHVNGQFLFALLRKAPCVTLHCPDRANVARTQSHVEVTLSEGETLGRVLA ADGTSALATACGVMDCEPYEQALAVIANVATSVHEGRAPERFOHQLIAPLPSMG RCSILVWCHLNEREVLMSDEKFCRELQSAFGKGLKITHAGKRSAYPLATITHARS ITHRTVLICNAQOTIHPITAGOSFNLGMDVNSLAETLTQAOBRCGEMDGVYLCAYOO RKQSORREITIGTDSLVHLFAHRNAPLVYVGRNIGLMTHELTTPADVLAQRTLGVAR " complement(7749. .9599) /note="similar to GenBank Accession Number D00398 (ECOPAP2)" complement(7906. .9231) /gene="pep" complement(7906. .9231) /gene="pep" /note="ORF_f441: third start codon" /codon_start=1 /transl_table=11 /protein_id="AA69076.1" /db_xref="GI:882438" /translation="MSEISRQEFORRRQALVEDMOPGSNALLFAAPEVTRSDSEPY RONSDFWPTGFNEPEAVLVLIKSDOTNHSVLPFNVRDLTAELWFGRIAGODAAPEK LGVDALAFSEINLOOLYOLNGLDVYVYHAGGEYAAADVIVNSALELKRGSNONLTAP ATMIDNRPPVHEMRPLKSPREIATVLRAGEITAMAKTRAMEKCRGMPFYLHKEGELHH EENRHQARTPSTNTIVSGENGCLHTTENECBMRDGLVLIDAGCEYGVACNDITRT FPVNGFTQAOEELIYDILESETSLRLTRPQTSILEVGEVVRIMVSLVKGLILKG DVDELAQNAHRPFEMHGLSHMLGDLVDVYGVGDORSRLPEGVLVLPVPGIYAVPD AEVPEYRCIGIRIEDVDIVITEGENLTAASVYKKRBEIEALMVARKO" complement(9257. .9841) /gene="ygfB" complement(9257. .9841) /note="ygfB" /codon_start=1 /transl_table=11 /protein_id="AA69077.1" /db_xref="GI:882439" /translation="MLMSIONEMPVYENNOXVLTNOOGTGLTPAEMHGLISGMICGND DSWLPDLHDLTNEGMAFGHEALQALRKHSATSADLDDGFLQLYLPDGDVSVFD BDALAGVNNHFLIGLGYOPKIDLTGEGEAIIDOLRNIAQLGYDEDEDEBELFMSL BEILEVRYAALLCHPTFTNHPQTPAPEVOKPTLH" 10003. .10332 /note="ORF_o109" /codon_start=1
misc_feature	CDS	complement(5370. .10317) /note="similar to GenBank Accession Number D90281 (ECOPYU)" complement(5506. .6708) /gene="v1sc" complement(5506. .6708) /gene="v1sc" /note="ORF_f400" /codon_start=1 /transl_table=11 /protein_id="AA69074.1" /db_xref="GI:882436" /translation="MQSVDAIVGGVAVGLAVACGCGSLRVAVLEORVPEPLANA PPOLRVSAINASEKLLRLGVMDILSRPASCYGVMEVMDKDSFGHISFDOSMGVS HUGHIVENSIVHYALMRKHAOSDITLLAPLELOVANGENETFLTLKDGSMILARLV IGADGNSVILNKADIPLTFMDYOHALLVATIRTEEPHDAAVARYHGEGLIAPLPS DPHICSYMSISPEAOBMOOASEDEFERANIALRDNELGCKVESAROVPLTGRVA RQFASHRLALVGDAHRTIHPLAGOSVNGPNDAAELIAELKRIHROGCDIOXYIARR YERSRHSAAALMLAGMGFRDLFGSTNPAAKLLRDLTGLADTLPGVAKPOLIRQAMGL NDLPEMLR" complement(6731. .7909) /gene="ubiH" complement(6731. .7909) /note="ubiH" /codon_start=1 /transl_table=11 /protein_id="AA69075.1" /db_xref="GI:882437" /translation="MSVIIIVGGAGATLALAIISPLSGALPVHLIETAVPESIAHPC EDGRAIALAAGCQQLARIGWQSLADCATATITVHVSDRGHACFTTAAEDYQALAL GGVVALHVNGQFLFALLRKAPCVTLHCPDRANVARTQSHVEVTLSEGETLGRVLA ADGTSALATACGVMDCEPYEQALAVIANVATSVHEGRAPERFOHQLIAPLPSMG RCSILVWCHLNEREVLMSDEKFCRELQSAFGKGLKITHAGKRSAYPLATITHARS ITHRTVLICNAQOTIHPITAGOSFNLGMDVNSLAETLTQAOBRCGEMDGVYLCAYOO RKQSORREITIGTDSLVHLFAHRNAPLVYVGRNIGLMTHELTTPADVLAQRTLGVAR " complement(7749. .9599) /note="similar to GenBank Accession Number D00398 (ECOPAP2)" complement(7906. .9231) /gene="pep" complement(7906. .9231) /gene="pep" /note="ORF_f441: third start codon" /codon_start=1 /transl_table=11 /protein_id="AA69076.1" /db_xref="GI:882438" /translation="MSEISRQEFORRRQALVEDMOPGSNALLFAAPEVTRSDSEPY RONSDFWPTGFNEPEAVLVLIKSDOTNHSVLPFNVRDLTAELWFGRIAGODAAPEK LGVDALAFSEINLOOLYOLNGLDVYVYHAGGEYAAADVIVNSALELKRGSNONLTAP ATMIDNRPPVHEMRPLKSPREIATVLRAGEITAMAKTRAMEKCRGMPFYLHKEGELHH EENRHQARTPSTNTIVSGENGCLHTTENECBMRDGLVLIDAGCEYGVACNDITRT FPVNGFTQAOEELIYDILESETSLRLTRPQTSILEVGEVVRIMVSLVKGLILKG DVDELAQNAHRPFEMHGLSHMLGDLVDVYGVGDORSRLPEGVLVLPVPGIYAVPD AEVPEYRCIGIRIEDVDIVITEGENLTAASVYKKRBEIEALMVARKO" complement(9257. .9841) /gene="ygfB" complement(9257. .9841) /note="ygfB" /codon_start=1 /transl_table=11 /protein_id="AA69077.1" /db_xref="GI:882439" /translation="MLMSIONEMPVYENNOXVLTNOOGTGLTPAEMHGLISGMICGND DSWLPDLHDLTNEGMAFGHEALQALRKHSATSADLDDGFLQLYLPDGDVSVFD BDALAGVNNHFLIGLGYOPKIDLTGEGEAIIDOLRNIAQLGYDEDEDEBELFMSL BEILEVRYAALLCHPTFTNHPQTPAPEVOKPTLH" 10003. .10332 /note="ORF_o109" /codon_start=1
		complement(5370. .10317) /note="similar to GenBank Accession Number D90281 (ECOPYU)" complement(5506. .6708) /gene="v1sc" complement(5506. .6708) /gene="v1sc" /note="ORF_f400" /codon_start=1 /transl_table=11 /protein_id="AA69074.1" /db_xref="GI:882436" /translation="MQSVDAIVGGVAVGLAVACGCGSLRVAVLEORVPEPLANA PPOLRVSAINASEKLLRLGVMDILSRPASCYGVMEVMDKDSFGHISFDOSMGVS HUGHIVENSIVHYALMRKHAOSDITLLAPLELOVANGENETFLTLKDGSMILARLV IGADGNSVILNKADIPLTFMDYOHALLVATIRTEEPHDAAVARYHGEGLIAPLPS DPHICSYMSISPEAOBMOOASEDEFERANIALRDNELGCKVESAROVPLTGRVA RQFASHRLALVGDAHRTIHPLAGOSVNGPNDAAELIAELKRIHROGCDIOXYIARR YERSRHSAAALMLAGMGFRDLFGSTNPAAKLLRDLTGLADTLPGVAKPOLIRQAMGL NDLPEMLR" complement(6731. .7909) /gene="ubiH" complement(6731. .7909) /note="ubiH" /codon_start=1 /transl_table=11 /protein_id="AA69075.1" /db_xref="GI:882437" /translation="MSVIIIVGGAGATLALAIISPLSGALPVHLIETAVPESIAHPC EDGRAIALAAGCQQLARIGWQSLADCATATITVHVSDRGHACFTTAAEDYQALAL GGVVALHVNGQFLFALLRKAPCVTLHCPDRANVARTQSHVEVTLSEGETLGRVLA ADGTSALATACGVMDCEPYEQALAVIANVATSVHEGRAPERFOHQLIAPLPSMG RCSILVWCHLNEREVLMSDEKFCRELQSAFGKGLKITHAGKRSAYPLATITHARS ITHRTVLICNAQOTIHPITAGOSFNLGMDVNSLAETLTQAOBRCGEMDGVYLCAYOO RKQSORREITIGTDSLVHLFAHRNAPLVYVGRNIGLMTHELTTPADVLAQRTLGVAR " complement(7749. .9599) /note="similar to GenBank Accession Number D00398 (ECOPAP2)" complement(7906. .9231) /gene="pep" complement(7906. .9231) /gene="pep" /note="ORF_f441: third start codon" /codon_start=1 /transl_table=11 /protein_id="AA69076.1" /db_xref="GI:882438" /translation="MSEISRQEFORRRQALVEDMOPGSNALLFAAPEVTRSDSEPY RONSDFWPTGFNEPEAVLVLIKSDOTNHSVLPFNVRDLTAELWFGRIAGODAAPEK LGVDALAFSEINLOOLYOLNGLDVYVYHAGGEYAAADVIVNSALELKRGSNONLTAP ATMIDNRPPVHEMRPLKSPREIATVLRAGEITAMAKTRAMEKCRGMPFYLHKEGELHH EENRHQARTPSTNTIVSGENGCLHTTENECBMRDGLVLIDAGCEYGVACNDITRT FPVNGFTQAOEELIYDILESETSLRLTRPQTSILEVGEVVRIMVSLVKGLILKG DVDELAQNAHRPFEMHGLSHMLGDLVDVYGVGDORSRLPEGVLVLPVPGIYAVPD AEVPEYRCIGIRIEDVDIVITEGENLTAASVYKKRBEIEALMVARKO" complement(9257. .9841) /gene="ygfB" complement(9257. .9841) /note="ygfB" /codon_start=1 /transl_table=11 /protein_id="AA69077.1" /db_xref="GI:882439" /translation="MLMSIONEMPVYENNOXVLTNOOGTGLTPAEMHGLISGMICGND DSWLPDLHDLTNEGMAFGHEALQALRKHSATSADLDDGFLQLYLPDGDVSVFD BDALAGVNNHFLIGLGYOPKIDLTGEGEAIIDOLRNIAQLGYDEDEDEBELFMSL BEILEVRYAALLCHPTFTNHPQTPAPEVOKPTLH" 10003. .10332 /note="ORF_o109" /codon_start=1


```

/transl_table=11
/protein_id="AAA69078.1"
/db_xref="GI:882440"
/translation="WSAQPVDIOIGRSLRVNCPDPDQDALNOAADLNOILDKER
TRVNTLEQVFLAALNISLEYELQEKAKTRDYASMEORIRMLQOTIEQALLEGRITE
KTNQNE"
misc_feature
10266..11219
/note="similar to GenBank Accession Number M12965
(ECOSSR)"
misc_feature
10374..10557
/note="similar to GenBank Accession Number X01238
(ECRRN6S)"
10374..10556
/feature="ssr"
10374..10556
/feature="ssr"
/misc_RNA
/product="stable 6S RNA"

Query Match      100.0% Score 1155; DB 1: Length 141744;
Best Local Similarity 100.0%; Pred. No. 1.4e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGAGAGGCGCATCTGACAAATTTGCT 60
    |||||||
Db 41101 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGAGAGGCGCATCTGACAAATTTGCT 41160

OY 61 GACCAATTTCTGATCCCGTTTACGCGCATCTCGAACAGAGATCCGAAGCAGCGCTT 120
    |||||||
Db 41161 GACCAATTTCTGATCCCGTTTACGCGCATCTCGAACAGAGATCCGAAGCAGCGCTT 41220

OY 121 GCTTGGAAACCTACGTAACAAACCGGCGATGTTTATGTTGGCGGCAATCACCACAGC 180
    |||||||
Db 41221 GCTTGGAAACCTACGTAACAAACCGGCGATGTTTATGTTGGCGGCAATCACCACAGC 41280

OY 181 GCCTGGTAGACATCGAAGAGATCACCCGTAACACCGTTCCGAATTTGGCTATGTGCAT 240
    |||||||
Db 41281 GCCTGGTAGACATCGAAGAGATCACCCGTAACACCGTTCCGAATTTGGCTATGTGCAT 41340

OY 241 TCCGACATGGGCTTTACGCTACCTCTGTCGGCTTCTGAGCGCTATCCGCAAAACAGTCT 300
    |||||||
Db 41341 TCCGACATGGGCTTTACGCTACCTCTGTCGGCTTCTGAGCGCTATCCGCAAAACAGTCT 41400

OY 301 CCTGACATCAACGAGGCGTTGACCGCTGCGCATCCCTCGAACAGCGCGGCGTGACACAG 360
    |||||||
Db 41401 CCTGACATCAACGAGGCGTTGACCGCTGCGCATCCCTCGAACAGCGCGGCGTGACACAG 41460

OY 361 GGTCTGATGTTGGCTACGCACTAATGAACCGAGCTGCTGATGCCAGCATTATCAC 420
    |||||||
Db 41461 GGTCTGATGTTGGCTACGCACTAATGAACCGAGCTGCTGATGCCAGCATTATCAC 41520

OY 421 TATGACACCGCTGTGACAGGCTGAGAGTGGTAAACACGCACTCTGCCCTGG 480
    |||||||
Db 41521 TATGACACCGCTGTGACAGGCTGAGAGTGGTAAACACGCACTCTGCCCTGG 41580

OY 481 CTGCGCCGAGCGAGAAACCGAGTACTTTTCAGTATGACGAGCAAAATCGTTGCT 540
    |||||||
Db 41581 CTGCGCCGAGCGAGAAACCGAGTACTTTTCAGTATGACGAGCAAAATCGTTGCT 41640

OY 541 ATGATGCTGCTGCTGCTTTCACCTCAGCACTCTGAGAGATTCGACGAAATGCTGCA 600
    |||||||
Db 41641 ATGATGCTGCTGCTGCTTTCACCTCAGCACTCTGAGAGATTCGACGAAATGCTGCA 41700

OY 601 GAAGCGTATGAGAGATCATCAAGCAATCTGCCGCTGAATGGAGCATCTTGGC 660
    |||||||
Db 41701 GAAGCGTATGAGAGATCATCAAGCAATCTGCCGCTGAATGGAGCATCTTGGC 41760

OY 661 ACCAAATTTCTCAACCCGAGCGTCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
    |||||||
Db 41761 ACCAAATTTCTCAACCCGAGCGTCTTTCGTTATCGGTGGCCCAATGGGTGACTGC 41820

OY 721 GGTCTGCTGCTGCTAAATTTATGTTGATACCTACGCGGCGATGCGGCTCAGGCTGC 780
    |||||||
Db 41821 GGTCTGCTGCTGCTAAATTTATGTTGATACCTACGCGGCGATGCGGCTCAGGCTGC 41880

```

```

OY 781 GGTGATCTCTGTTAAGATCATCAAAAGTGGAGCGTTCCGACCATCAGGAGCGT 840
    |||||||
Db 41881 GGTGATCTCTGTTAAGATCATCAAAAGTGGAGCGTTCCGACCATCAGGAGCGT 41940

OY 841 TATGTCGCAAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
    |||||||
Db 41941 TATGTCGCAAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42000

OY 901 TACGCAATCGGCGCTGCTGACACCGATCCATCATGTGTGAATCTTCCGCTACTGAGAA 960
    |||||||
Db 42001 TACGCAATCGGCGCTGCTGACACCGATCCATCATGTGTGAATCTTCCGCTACTGAGAA 42060

OY 961 GTCCCTTCTGAACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
    |||||||
Db 42061 GTCCCTTCTGAACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42120

OY 1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
    |||||||
Db 42121 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42180

OY 1081 TTTGTCGTGACATTTTCCCGTGGGAAACACCGAAGCGAGCTGCTGCGCATGCT 1140
    |||||||
Db 42181 TTTGTCGTGACATTTTCCCGTGGGAAACACCGAAGCGAGCTGCTGCGCATGCT 42240

OY 1141 GCCGCTCGAAGTAA 1155
    |||||||
Db 42241 GCCGCTCGAAGTAA 42255

RESULT 10
AE005524 11178 bp DNA linear HCT 21-MR-2001
LOCUS AE005524
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 143
ACCESSION AE005524
VERSION AE005524.1
KEYWORDS GI:12517476
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae: Escherichia

REFERENCE
1 (bases 1 to 11178)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
21074935
11206551
2 (bases 1 to 11178)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES
source
1..11178
Location/Qualifiers
1..11178
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/feature="enterohemorrhagic"
complement(188..2179)
/feature="tkA"
/feature="24279"

```

```

CDS      complement(188..2179)
        /gene="lta"
        /function="enzyme: Central intermediary metabolism:
        Non-oxidative branch, pentose py"
        /note="Residues 1 to 663 of 663 are 99.69 pct identical to
        residues 1 to 663 of 663 from Escherichia coli K-12 strain
        MG1655: B2935"
        /codon_start=1
        /transl_table=11
        /product="transketolase 1 isozyme"
        /protein_id="AAG58065.1"
        /db_xref="GI:12517477"
        /translation="MSSRRELNANIRALSMDAVQAKSGHPGPMGMADIAEVLNRD
        LKHNPNQSWADRRPRLVSLNGHSMILYSLHGTGDIPELKNRQLSKTPGHE
        VGTACVETTPICGCIANAYGMALAEKTLAOKNPGCHDIVHTYAFMGDCGME
        GISHYCSLAGTLKGLTAIFDDNGISIDGVECHFTDDTAMREAYCNHTRIDIG
        HDASIKRAVEARAVTDKPSLCKKTIIGFSPNAGTHDSGAPLQDAETALIREO
        LGMKYAPEIPEIYIOMDAKESQAKESAMKEFAAYAKAQEAETTRMKDEP
        SPDKAKKEFIKLOANPAKIASRKASQNAIEAFGLPEPLGSDAPLSMLTWSG
        SKAINEDAGNYIHGVEFGMTAINGFSLHGFLPYSTFLMEVEYARNVMAAL
        MKORVMYTHDSIGLEDGPTHOPEOVAASLRVTPNMSTMRPCDQVESAVAMKQVE
        RODGPALILSRONLAODRTEOLANIRGCVLKDCAQOPELLFIATGSEVELAVA
        AYEKLAEGVKARVVSMPSTDAFDKODAVRESVLKAVTAHAAVAGIADYWKYVG
        LNCALIGMTTFGESAPAEILFEFGFTVDNVAKAKELL"
        /gene="y9gC"
        /note="24280"
        /db_xref="GI:12517481"
        /function="orf: Unknown function"
        /note="Residues 1 to 294 of 294 are 99.31 pct identical to
        residues 1 to 294 of 294 from Escherichia coli K-12 strain
        MG1655: B2936"
        /codon_start=1
        /transl_table=11
        /product="orf, hypothetical protein"
        /protein_id="AAG58066.1"
        /db_xref="GI:12517478"
        /translation="MLACCAIYSPVRCAPPYTEREPTCKRIFRILITLGRKMK
        ITRALLVMSVATVLTGCONMDSNGLSSGEAQAQSLSDAOKTISDOACQMDKMA
        T1APANSEYAKRLTTIANLGNINQOPVNYKVYAKNDVAEMANGCIRKIVSGLDL
        MTNVEEAVYGHENGVHVGKGMQVALGTNAVVAASAGIVGSSLSQSGDLG
        EKLNSOESOROEADVDYDLRLRGRISPAQLTSFEKLAKLEGRROSSMDDHPA
        SARARHIDRMSADGIR"
        /complement(3421..4341)
        /gene="speB"
        /note="24281"
        /complement(3421..4341)
        /gene="speB"
        /function="enzyme: Central intermediary metabolism:
        Polyamine biosynthesis"
        /note="Residues 1 to 306 of 306 are 100.00 pct identical
        to residues 1 to 306 of 306 from Escherichia coli K-12
        strain MG1655: B2937"
        /codon_start=1
        /transl_table=11
        /product="agmatinase"
        /protein_id="AAG58067.1"
        /db_xref="GI:12517479"
        /translation="MSTLGHQYDMSLVNAFGRLRLPMPQYDSADAWITGVPEDM
        ATSGRACGRHPAIAROYSTNLAEHNRFPNEMRRLNVDCGLVYFGDAREMS
        EKLQAHAEKLAGKRLMSFGDHFVLPLLRAHAKHGMALVHPDAHDTYANGCE
        FDKHTMYTAPKEGLIDPNHSVOIGITFEEDKDSGFTVLDACOVNDRSVDDVIAOVQ
        IYCDMPVYLTEDIDCDLPAFGTGTGTVIGGLTSRAIKVRLGLKLDNIYGMVDEYA
        PAYDOSEITLAAATLALMLYIOAKRKE"
        /note="O-island #120: Region of the EDL933 chromosome not
        homologous to E. coli K-12 MG1655"
        /complement(4477..5208)
        /gene="24282"
        /complement(4477..5208)
        /gene="24282"
        /function="orf: Unknown function"

gene      /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /protein_id="AAG58068.1"
        /db_xref="GI:12517480"
        /translation="MKKKVYSALVALIYLLAGSSNMYNSASGNYGTAMGQVHS
        TVQGSARARARNDPEAKIVISTHVPSCYDRVYSIRINLEIYAIRDNFNSLPTRY
        YDSSNNEPRYIYHARVGMNYOLYVNNYSRNTYETIYAVTDGLVNGQGSNNNGYI
        VNAGSLAIKGFRRKDKHTEAFOFANVADSVAAANSAGDVRNTGVYIQAPELQGOAO
        NALPPCSGAOFADPNNSYAPPCRK"
        /complement(5354..7330)
        /gene="speA"
        /note="24283"
        /complement(5354..7330)
        /gene="speA"
        /function="enzyme: Central intermediary metabolism:
        polyamine biosynthesis"
        /note="Residues 1 to 658 of 658 are 99.84 pct identical to
        residues 1 to 658 of 658 from Escherichia coli K-12 strain
        MG1655: B2938"
        /codon_start=1
        /transl_table=11
        /product="biosynthetic arginine decarboxylase"
        /protein_id="AAG58069.1"
        /db_xref="GI:12517481"
        /translation="MSDDSMGLPSSAGEHYLRSNOEYAMSSQENSKRLRTYNIAM
        GNNYDVNDELCHISVCPDPDVEARVDLAOLVKTREAOGRPALFCFPOIIOHRLS
        INAKFRARESYGNDYDFLVYPIKNOHRYVESLIHSGEPLGLEAGKAEIMAVPA
        HGMTRSVYICNGYKRDREYIRLALIGEMKHVYLVIEKSELAIVLDAEAFRLNVP
        LQVRRARLASQSGSKNOQSSGCEKSKGLATOVOLVETLREGRDLSIOLHPHLSO
        MANIRDIATGVESEARFVELKHLKLVNIQCFVGGGLGVDEGTSOSQSYNYGLNE
        VANNIIMAGDACEENGILPPIYVITESGAIVAHIVLNSITIGVERNYPVTPAAE
        DAPRALOSMETHNOEMIEPTGRSLRESADMDLHDHITIGSSCTVSLQRAVAEO
        LYLSMCHVOXKOLDPONRAHPIIDELQERMDKMYNVSLSFQSMIDANGIIDLPLVL
        PLQGLDOVERPARRVLIIDICDSDGATDHIIDDDGATTPMPMEYDENPMIGFPPVG
        AVQELIGNMHNLRGDTNAVDFYFPDGSFEVLESDSGDGVADMLOYVOLDPHTLLTGF
        RQVKRTDLDAELQQQVLEEFEXGLGYLYLDE"
        /complement(7339..7485)
        /gene="y9gB"
        /note="24284"
        /complement(7339..7485)
        /gene="y9gB"
        /function="orf: Unknown function"
        /note="Residues 1 to 48 of 48 are 100.00 pct identical to
        residues 1 to 48 of 48 from Escherichia coli K-12 strain
        MG1655: B2939"
        /codon_start=1
        /transl_table=11
        /product="orf, hypothetical protein"
        /protein_id="AAG58070.1"
        /db_xref="GI:12517482"
        /translation="MLHNVMKKKPVQALEQHSLSLENPCAYGILSOFQAAIVNCFTL
        NKRII"
        /note="7821"
        /gene="y9gC"
        /note="24285"
        /note="7821"
        /gene="y9gC"
        /function="orf: Unknown function"
        /note="Residues 1 to 71 of 71 are 100.00 pct identical to
        residues 1 to 71 of 71 from Escherichia coli K-12 strain
        MG1655: B2940"
        /codon_start=1
        /transl_table=11
        /product="orf, hypothetical protein"
        /protein_id="AAG58071.1"
        /db_xref="GI:12517483"
        /translation="MGITSAGMOSRDAECGERJFTRTVOVKQOTVHYVSPRPV
        KTNPOAKTILISTRIEYVTRKKRVLFI"
        /complement(7818..8069)
        /gene="y9gD"
        /note="24286"
        /complement(7818..8069)

```

```

/gene="ygd"
/function="orf: unknown function"
/notes="Residues 1 to 83 of 83 are 98.79 pct identical to
residues 1 to 83 of 83 from Escherichia coli K-12 strain
MG1655: B2941"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AG58072.1"
/db_xref="GI:12517484"
/translation="MVADPTTLOVKNTGSLSVRYGINIMAILCOFTTRPFLPEE
SCLILKTLWEIFPDNGLIRLYLLOPSAIVGYTRRA"
8125..9279

```

Query Match 98.9% Score 1142.2; DB 1; Length 11178;

Best local similarity 99.3% *Pred. No. 1.5e-263;

Matches 1147; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

Oy 1 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGGCATCTGACAAATTGCT 60
Db 8125 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGGCATCTGACAAATTGCT 8184
Oy 61 GACCAAAATTTCTGATGCCGTTTAAAGCGGATCTCGAAGAGATCCGAAGCAGCGTT 120
Db 8185 GACCAAAATTTCTGATGCCGTTTAAAGCGGATCTCGAAGAGATCCGAAGCAGCGTT 8244
Oy 121 GCTTGGCAAACTAGCTAAAAACGGGATGTTTATGTTGGGGGGAATCACCACG 180
Db 8245 GCTTGGCAAACTAGCTAAAAACGGGATGTTTATGTTGGGGGGAATCACCACG 8304
Oy 181 GCCTGGTAGACATCGAAGAGATCACCCGTAACACCGTTCCGGAATTGGCTATGCT 240
Db 8305 GCCTGGTAGACATCGAAGAGATCACCCGTAACACCGTTCCGGAATTGGCTATGCT 8364
Oy 241 TCCGACATGGGCTTTGACGCTTACTCTGTGGGTTCTGAGCGCTATGCGCAACAGT 300
Db 8365 TCCGACATGGGCTTTGACGCTTACTCTGTGGGTTCTGAGCGCTATGCGCAACAGT 8424
Oy 301 CCTGACATCAACACGAGCGGTTGACCGTGCATCCGCTGGGAAGAGCGCGGTCAGC 360
Db 8425 CCTGACATCAACACGAGCGGTTGACCGTGCATCCGCTGGGAAGAGCGCGGTCAGC 8484
Oy 361 GCTGTGATGTTTGGCTAGCACTAATGAACGAGCTGATGCCACACCTATAC 420
Db 8485 GCTGTGATGTTTGGCTAGCACTAATGAACGAGCTGATGCCACACCTATAC 8544
Oy 421 TATGACACACCGTCTGCTACAGCGCTGAAGTGCCTGAAGAGGCACTTGCCTG 480
Db 8545 TATGACACACCGTCTGCTACAGCGCTGAAGTGCCTGAAGAGGCACTTGCCTG 8604
Oy 481 CTGGCGCGGAGCGCAAAAGCAGGTAATTTCAATGATGACGAGCGCAAAATCGTTG 540
Db 8605 CTGGCGCGGAGCGCAAAAGCAGGTAATTTCAATGATGACGAGCGCAAAATCGTTG 8664
Oy 541 ATGATGATGCTGTGCTTTCACACTCAGCACTCGAAGAGATCGACAGAATCGCTG 600
Db 8665 ATGATGATGCTGTGCTTTCACACTCAGCACTCGAAGAGATCGACAGAATCGCTG 8724
Oy 601 GAAGCGGTAATGAAGAGATCATCAAGCAATTTGCGCGCTGAATGCTGACTTGC 660
Db 8725 GAAGCGGTAATGAAGAGATCATCAAGCAATTTGCGCGCTGAATGCTGACTTGC 8784
Oy 661 ACCAAATTTCTTCAACACCGGCTGCTTTCGTTATCGGTCGCAATGGTGACTGC 720
Db 8785 ACCAAATTTCTTCAACACCGGCTGCTTTCGTTATCGGTCGCAATGGTGACTGC 8844
Oy 721 GGTCTGACTGTCGTAATAATATACGTTATACACGCGGATGCGCGTACAGGTGC 780
Db 8845 GGTCTGACTGTCGTAATAATATACGTTATACACGCGGATGCGCGTACAGGTGC 8904
Oy 781 GGTGATTCCTGCTGAAGATCCATCAAAAGTGCAGCGTTCCGACACCTACGACGCT 840
Db 8905 GGTGATTCCTGCTGAAGATCCATCAAAAGTGCAGCGTTCCGACACCTACGACGCT 8964

```

```

Oy 841 TATGCGGCAAAACATGCTGCTGCTGCGGCGGATCTGTTGAATTCAGCTTCC 900
Db 8965 TATGCGGCAAAACATGCTGCTGCTGCGGCGGATCTGTTGAATTCAGCTTCC 9024
Oy 901 TACGCAATCGCGTGGCTGAACCGACGTCATCATGTGAGAACTTTGCTACTGAGAA 960
Db 9025 TACGCAATCGCGTGGCTGAACCGACGTCATCATGTGAGAACTTTGCTACTGAGAA 9084
Oy 961 GTGCTTTGTAACACATACCTGCTGCTAGCTGAGTTCTTTCGACCTCGCCCATACG 1020
Db 9085 GTGCTTTGTAACACATACCTGCTGCTAGCTGAGTTCTTTCGACCTCGCCCATACG 9144
Oy 1021 CTGATTACATGCTGATCTGCTGTCACCGATCTCAAGAAACCGACATACGCTAC 1080
Db 9145 CTGATTACATGCTGATCTGCTGTCACCGATCTCAAGAAACCGACATACGCTAC 9204
Oy 1081 TTTGCTGTGAACATTTTCCCGTGGGAAAAAACCGCAAGCGCAGCTCTGCGCATGCT 1140
Db 9205 TTTGCTGTGAACATTTTCCCGTGGGAAAAAACCGCAAGCGCAGCTCTGCGCATGCT 9264
Oy 1141 GCCGCTGTAAGTAA 1155
Db 9265 GCCGCTGTAAGTAA 9279

```

RESULT 11

```

AP002563 26658 bp DNA linear BCF 07-MAR-2001
LOCUS Escherichia coli O157:H7 complete genome, section 14/20.
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 14/20.
ACCESSION AP002563 BA000007
VERSION AP002563.1 GI:13363121
KEYWORDS
SOURCE

```

ORGANISM

```

DNA.
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae:
Escherichia.

```

REFERENCE

```

1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780

```

JOURNAL

```

MEDLINE
REFERENCE
AUTHORS
TITLE
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356

```

JOURNAL

```

MEDLINE
REFERENCE
AUTHORS
TITLE
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182

```

```

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
4 (sites)
Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsuno, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kihara, S., Shiba, T., Hattori, M. and
Shinagawa, H.

```


gene
CDS

SLRGKSIENKLNRIYALFMLVTTIVATLNPILQMIETLGPILIMILFLMPY
AIOKPAHKKYSGHISNVEVVMGLAISAFYSIFS*
7285: 8652
/gene="EC33657"
7285: 8652
/gene="EC33657"
/note="similar to SDAA_ECOLI q1178116 percent identity
77 in 455 aa, similar to SDBA_ECOLI q1178161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"

Query Match 98.9% Score 1142.2; DB 1: Length 26658;
Best Local Similarity 99.3%; Pred. No. 1.7e-263;
Matches 1147: Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 ATGGCAAAACACCTTTTACGTCGCGAGTCGCTCTGAAGGCGATCTTACAAATTCGT 60
DB 183501 ATGGCAAAACACCTTTTACGTCGCGAGTCGCTCTGAAGGCGATCTTACAAATTCGT 183560
OY 61 GACCAAAATTTCTGATGCGCTTTTACGCGATCTCTGAAGGCGATCTTACAAATTCGT 120
DB 183561 GACCAAAATTTCTGATGCGCTTTTACGCGATCTCTGAAGGCGATCTTACAAATTCGT 183620
OY 121 GCTTGGCAAACTTACGTCGCGAGTCGCTCTGAAGGCGATCTTACAAATTCGT 180
DB 183621 GCTTGGCAAACTTACGTCGCGAGTCGCTCTGAAGGCGATCTTACAAATTCGT 183680
OY 181 GCGTGGTACATGACGATGACGATGACGATGACGATGACGATGACGATGACGAT 240
DB 183681 GCGTGGTACATGACGATGACGATGACGATGACGATGACGATGACGATGACGAT 183740
OY 241 TCGGACATGCGCTTTGACGCTTACGCTTCTGCGCTTCTGAGCGCTATCGGCAACAGCT 300
DB 183741 TCGGACATGCGCTTTGACGCTTACGCTTCTGCGCTTCTGAGCGCTATCGGCAACAGCT 183800
OY 301 CCTGACATACACGAGCGCTTACGCTTCTGCGCTTCTGAGCGCTATCGGCAACAGCT 360
DB 183801 CCTGACATACACGAGCGCTTACGCTTCTGCGCTTCTGAGCGCTATCGGCAACAGCT 183860
OY 361 GGTCTGATGTTGGTACGCAACTATGAACGAGCGCTGCTGTCGACAGCTATCACC 420
DB 183861 GGTCTGATGTTGGTACGCAACTATGAACGAGCGCTGCTGTCGACAGCTATCACC 183920
OY 421 TATGCAACACGCTTCTGACAGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 480
DB 183921 TATGCAACACGCTTCTGACAGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 183980
OY 481 CTGCGCCGCGAGCGCAAAAGCCAGGTGACTTTTCACTATGACGAGCGCAAAATCGTTGCT 540
DB 183981 CTGCGCCGCGAGCGCAAAAGCCAGGTGACTTTTCACTATGACGAGCGCAAAATCGTTGCT 184040
OY 541 ATGAGATGCTGCTGCTTCTCACTACGACTGAGAGATGAGAGATGAGAGATGAGAGAT 600
DB 184041 ATGAGATGCTGCTGCTTCTCACTACGACTGAGAGATGAGAGATGAGAGATGAGAGAT 184100
OY 601 GAAGCGGTATGAGAGATGATCAAGCAATTCCTCCGCTGAATGCGTACTTCTGCC 660
DB 184101 GAAGCGGTATGAGAGATGATCAAGCAATTCCTCCGCTGAATGCGTACTTCTGCC 184160
OY 661 ACCAAATTTCTTATCAACCGGAGCGTCTTCTGATGCGGTGCGCCATGGTGAAGTGC 720
DB 184161 ACCAAATTTCTTATCAACCGGAGCGTCTTCTGATGCGGTGCGCCATGGTGAAGTGC 184220
OY 721 GGTCTACTAGTGTGTAATATGCTTACTAGCGGCGCATGGCGCTCTACGCTGGC 780
DB 184221 GGTCTACTAGTGTGTAATATGCTTACTAGCGGCGCATGGCGCTCTACGCTGGC 184280
OY 781 GGTGCAATTTCTGTAAGATGATCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 840
DB 184281 GGTGCAATTTCTGTAAGATGATCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 184340
OY 841 TATGTCGCAAAACATCTGCTGCTGCGCTGCGAGCTGTTGAATTCAGGTTTCC 900
DB 184341 TATGTCGCAAAACATCTGCTGCTGCGCTGCGAGCTGTTGAATTCAGGTTTCC 184400

OY 901 TACGCAATGCGGCGCTGACGAGCTTCCATCATGTGTAAGTCTTGGTACTGAGAA 960
DB 184401 TACGCAATGCGGCGCTGACGAGCTTCCATCATGTGTAAGTCTTGGTACTGAGAA 184460
OY 961 GTGCTTCTGAACACGAGCTTCTGATGAGTCTTCTGAGCTTCTGAGCTTCTGAGCT 1020
DB 184461 GTGCTTCTGAACACGAGCTTCTGATGAGTCTTCTGAGCTTCTGAGCTTCTGAGCT 184520
OY 1021 CTGATTCACATGCTGATGCTGCTGACACCGATCTACAAAGAACGCGACATACGCTAC 1080
DB 184521 CTGATTCACATGCTGATGCTGCTGACACCGATCTACAAAGAACGCGACATACGCTAC 184580
OY 1081 TTTGGTGTGAACATTTCCGCTGGGAAACGAGCAAGCGAGCTGCTGGCGCATGCT 1140
DB 184581 TTTGGTGTGAACATTTCCGCTGGGAAACGAGCAAGCGAGCTGCTGGCGCATGCT 184640
OY 1141 GCCGCTCTGAAGTAA 1155
DB 184641 GCCGCTCTGAAGTAA 184655

RESULT 12

ECOMETX 1152 bp DNA linear BCT 31-JUL-1996
ECOMETX
LOCUS
DEFINITION
Escherichia coli S-adenosylmethionine synthetase II dimer subunit
(metx) gene, complete cds.
VERSION
M98266.1 GI:146850
KEYWORDS
S-adenosylmethionine synthetase II; metx gene.
SOURCE
Escherichia coli (strain K-12) (library: Lambda) DNA.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 1152)
AUTHORS
Satschidan, C., Taylor, J. C. and Markham, G. D.
TITLE
Isoczymes of S-adenosylmethionine synthetase are encoded by tandemly
duplicated genes in Escherichia coli
JOURNAL
Mol. Microbiol. 9 (4), 835-846 (1993)
MEDLINE
94049123
PubMed
8231813

FEATURES

source
1..1152
Location/Qualifiers
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/issue_1="lambda"
1..1152
/gene="metx"
1..1152
/gene="metx"
/EC_number="2.5.1.6"
/codon_start=1
/transl_table=11
/product="S-adenosylmethionine synthetase II"
/protein_id="AAB05197.1"
/db_xref="GI:146851"

translation="MAKHLFSESSEGHPOKIAOISDAVLDALBEDPKARVACET
YKVTGMLVGEITTSVAVDIEITRNVRREIGVHSDMGFPANSCAVLSAIGKOSP
INOGVDRADPLEOGAGDOGLMGVATNEDVLMPIYAHNLVORAEVRANKTILRV
RPDASOYTFQVDDKIGVLDVAVLSTGHSSEITDOKSLOEAVMEI1KPI1LPAEMVTS
ATKFINPTGRFVIGGPMGDCGLTGRTKLIIVDYCGMARHGGASFGKSPSKVRSAAV
ARKYAKNIIVANGLDRCIEIOVSYAIGCAEPISINVERFTGTEKVSBDLTLLVREFPD
LNPGLIOMLDLHPHYETAYVGHFHEHPWEXTDRAHVLPDAAGLK"

misc_difference
/gene="metx"
/note="different from metx"

misc_difference
123..135
/gene="metx"
/note="different form metx"

BASE COUNT 277 a 310 c 302 g 263 t
ORIGIN

Query Match 97.8% Score 1129.4; DB 1: Length 1152;
 Best Local Similarity 99.2%; Pred. No. 1.6e-260;
 Matches 1146; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

1 ATGGCAAAACACCTTTTACCTCCGAGTCCCTCTCTGAAGGCACTCCAGCAAAATTCCT 60
 1 ATGGCAAAACACCTTTTACCTCCGAGTCCCTCTCTGAAGGCACTCCAGCAAAATTCCT 60
 61 GACCAAAATTTCTGATGCCGTTTAAAGCCGATCCTCGAACAGATCCGAAGCACGCCGT 120
 61 GACCAAAATTTCTGATGCCGTTTAAAGCCGATCCTCGAACAGATCCGAAGCACGCCGT 120
 121 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 180
 121 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 180
 181 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 240
 181 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 240
 241 TCCGACATGGGCTTTGACGCTACTCTGCGGTTCTGAGCGCTATCGCAACAGCTCT 300
 241 TCCGACATGGGCTTTGACGCTACTCTGCGGTTCTGAGCGCTATCGCAACAGCTCT 300
 301 CCTGACATCAACGAGGGGCTTACGCTGCGATCCGCTGGAACAGGGCGGGGAGACAG 360
 301 CCTGACATCAACGAGGGGCTTACGCTGCGATCCGCTGGAACAGGGCGGGGAGACAG 360
 361 GGTCTGATGTTTGGCTACGCACTAATGAAGACGAGCTGTGATGCCAGCACTATCAC 420
 361 GGTCTGATGTTTGGCTACGCACTAATGAAGACGAGCTGTGATGCCAGCACTATCAC 420
 421 TATGACACCGCTCTGCTGATACAGCTGAGCTGAGCTTAAAGAGGCACTCTGCGTGG 480
 421 TATGACACCGCTCTGCTGATACAGCTGAGCTGAGCTTAAAGAGGCACTCTGCGTGG 480
 481 CTGGCCCGGAGCGGCAAAAGCAGGCTGATTTTCACTGATGACAGCGGCAAAATCGTTGT 540
 481 CTGGCCCGGAGCGGCAAAAGCAGGCTGATTTTCACTGATGACAGCGGCAAAATCGTTGT 540
 541 ATGCAATGCTGCTGCTGCTTCCACTCAGCATCTGTAAGAGATGACACAGCAAAATCGTCCA 600
 541 ATGCAATGCTGCTGCTGCTTCCACTCAGCATCTGTAAGAGATGACACAGCAAAATCGTCCA 600
 601 GAACCGGTAATGGAAGAGATCATCAAGCCATTTCCGCCGCTGAATGGCTGACTTCTGCC 660
 601 GAACCGGTAATGGAAGAGATCATCAAGCCATTTCCGCCGCTGAATGGCTGACTTCTGCC 660
 661 ACCAAATCTTTCATCAACCCGAGCGGCTGTTTCCGTTACGTTGGGCGCAATGGTGACTGC 720
 661 ACCAAATCTTTCATCAACCCGAGCGGCTGTTTCCGTTACGTTGGGCGCAATGGTGACTGC 720
 721 GGTCTGACTGCTGTAATAATTATCGTTGATCACTACGCGGCGCATGCGCGTACAGCTGGC 780
 721 GGTCTGACTGCTGTAATAATTATCGTTGATCACTACGCGGCGCATGCGCGTACAGCTGGC 780
 781 GGTCTGACTGCTGTAATAATTATCGTTGATCACTACGCGGCGCATGCGCGTACAGCTGGC 840
 781 GGTCTGACTGCTGTAATAATTATCGTTGATCACTACGCGGCGCATGCGCGTACAGCTGGC 840
 841 TATGCGGCAAAACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 841 TATGCGGCAAAACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 901 TACGCAATGGCGTGGTGAAGACGAGCTCATCATGTAGAAACTTTGGTACTGAGAAA 960
 901 TACGCAATGGCGTGGTGAAGACGAGCTCATCATGTAGAAACTTTGGTACTGAGAAA 960
 961 GTGCTTCTGAAACACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 961 GTGCTTCTGAAACACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080

1018 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
 1081 TTTGGTCTGTAACATTTTCCCTGGGAAATAACGCAACAGGCAAGCTGCTGCGGATGCT 1140
 1078 TTTGGTCTGTAACATTTTCCCTGGGAAATAACGCAACAGGCAAGCTGCTGCGGATGCT 1137
 1141 GCCGCTCTGAAGTAA 1155
 1138 GCCGCTCTGAAGTAA 1152

RESULT 13
 ECOMETK
 LOCUS ECOMETK 1462 bp DNA linear BCT 26-Apr-1993
 DEFINITION E. coli metK gene coding for S-adenosylmethionine synthetase.
 ACCESSION K02129
 VERSION K02129.1 GI:146838
 KEYWORDS S-adenosylmethionine synthetase; metK gene; synthetase.
 SOURCE E. coli (K12) DNA, clone pK8.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
 AUTHORS Markham G.D., Deparasis J. and Galimaitan J.
 TITLE The sequence of metK, the structural gene for S-adenosylmethionine synthetase in Escherichia coli
 JOURNAL J. Biol. Chem. 259 (23), 14505-14507 (1984)
 MEDLINE 85054924
 PUBMED 6094561

COMMENT
 Sequence furnished by Markham (07-APR-1985) on floppy disk. [1]
 notes a potential ribosome-binding site at 72-77.

FEATURES
 source
 1..1462
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 gene
 86..1240
 /gene="metK"
 CDS
 86..1240
 /gene="metK"
 /codon_start=1
 /transl_table=1
 /product="S-adenosylmethionine synthetase"
 /protein_id="AA024164.1"
 /db_xref="GI:146838"
 /translation="MASHLPTSESYSVSGHFDKIDQISDAVLDAILEODPKARYACT
 YVKTGICFSMRHHQIPWVDIEITRNTYREIGYVSHDGFANSCAVLSAICKOSP
 DIRQGVDRADPLEGADQGLDVSATLPTCLMPAPITYAHRLVORQAEVRKNGTL
 RVRPDAKSQVTFSDYDKIVGIDAVVLSHSEEDIDKSLQEAAMEEIRKPIPAEVL
 TSAKTFEINPTGFRVIGGPMGDCGLTGKRIIVDTGGMHRRGAFSGKDPKVDRA
 AYAKRYAKNIYAAGLADRCIEIOYSAIGLABPRTINVERFQTEKVPSEQLTLVREF
 FDLPIGLIOMLDLHPILYKETAATGHEGRHFWKTDKQVLDLDAAGLK"

BASE COUNT
 368 a 389 c 349 g 356 t
 ORIGIN 5 bp upstream of PstI site.

Query Match 90.3% Score 1043.4; DB 1: Length 1462;
 Best Local Similarity 97.2%; Pred. No. 7.4e-240;
 Matches 1128; Conservative 0; Mismatches 21; Indels 12; Gaps 6;

1 ATGGCAAAACACCTTTTACCTCCGAGTCCCTCTCTGAAGGCACTCCAGCAAAATTCCT 60
 86 ATGGCAAAACACCTTTTACCTCCGAGTCCCTCTCTGAAGGCACTCCAGCAAAATTCCT 145
 61 GACCAAAATTTCTGATGCCGTTTAAAGCCGATCCTCGAACAGATCCGAAGCACGCCGT 120
 146 GACCAAAATTTCTGATGCCGTTTAAAGCCGATCCTCGAACAGATCCGAAGCACGCCGT 205
 121 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 179
 206 GCTTGGCAAACTACGTAAGAAACCGGCAATGTTTAAAGTGGGCAATACCAACCCAGC 265
 180 CG--CTTGGTAGACATCGAAGAGATCACCCGTAACACCGTTCCGAATTTGGCTATGCTG 237

VERSION	KEYWORDS
AEO08842.1	GI:16421636
SOURCE	ORGANISM
Salmonella typhimurium LT2.	Salmonella typhimurium LT2
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	Salmonella.
REFERENCE	AUTHORS
1 (bases 1 to 22204)	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Portwolk, S., All, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W., Stenelev, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium LT2	
JOURNAL	MEDLINE
Nature	413 (6858), 852-855 (2001)
PUBMED	REFERENCE
21534948	11677609
2 (bases 1 to 22204)	The Salmonella typhimurium Genome Sequencing Project.
AUTHORS	TITLE
Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA	
JOURNAL	COMMENT
Supported by NIH grant 5U 01 AI43283	
Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GIMMER: http://www.tigr.org/software/glimmer/ and Genemark: http://opal.biology.gatech.edu/genemark/	
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes: http://www.genome.ad.jp/kegg/ , and Pedro Romero and Peter Karp at Ecocyc: http://ecocyc.org/ecocyc/	
The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB: http://kinich.cfin.unam.mx:8850/db/reguondb_intro.frameset	
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.	
FEATURES	SOURCE
Location/Qualifiers	
1..22204	/organism="Salmonella typhimurium LT2"
/strain="LT2: SGSC 1412; ATCC:700720"	
/db_xref="ATCC:700720"	
/db_xref="taxon:99287"	
/note="LT2"	
complement(657..1463)	
/gene="STM3084"	
complement(657..1457)	
/gene="STM3084"	
/note="similar to E. coli regulator for uxu operon (AAC77280.1): Blast hit to AAC77280.1 (257 aa), 39%	
identity in aa 8 - 247"	
/codon_start=1	
/transl_table=1	
/product="putative gnr family regulatory protein"	
/protein_id="AAL21959.1"	
/db_xref="GI:16421637"	
/translation="MPTLPSSGSAIVYICMGKQIILTKRRFDIGLOIEELYSVGRKAGELEPAERELSERFOTSRITIKETIIMLEKGVEVKQGAGTVFTOSLERINOKAILLPSYDIPGPELQAROVLESNITGEAAOIRFNELKOLKRIILEDEKOICGSODPFEEIDROHAKTIAESTONRVLYIKOASAELEMAVRVPRMKOLNVKYLLRKELRMKWEDHRSSIFLAOLKRDAEAOASATHLENKNELVIIFQDDSLIEDFDFFPAT"	
complement(1458..1463)	
/gene="STM3084"	
/note="putative RBS for STM3084; ReguonDB: STMS1H003125"	

gene complement(1611. .2348)
/gene="STM3085"
complement(1611. .2342)
/gene="STM3085"
/codon_start=1
/transl_table=11
/product="putative outer membrane lipoprotein"
/protein_id="AAL21960.1"
/db_xref="GI:16421638"
/translation="MSKPKFNAIVIGVFLLLGCGSSCMQSNSSSDVGTAMGDVHSSVQSVYRARRRREPAAIINIVSTQYPSGDVYSIISDLEVARPAPANSPIPIIRVNASGCMQYISIPARGCKNNVOLIIRNYSHDNVEIVATVQDLVNLGKASLNHHGIVNAGSLAIKFRKDKHTEARFCFADVAADVAHNSAGDVRLNICVIGFAFALQGCATNTLPSCQAFPADNNGTAPPCRK"
complement(2343. .2348)
/gene="STM3085"
/note="putative RBS for STM3085: RegulonDB: STMS1H003126"
complement(2486. .4558)
/gene="spea"
/note="spea"
/note="spea"
/EC_number="4.1.1.19"
/note="similar to E. coli biosynthetic arginine decarboxylase (AAC75975.1): Blastp hit to AAC75975.1 (658 aa), 97% identity in aa 1 - 658"
/codon_start=1
/transl_table=11
/product="arginine decarboxylase"
/protein_id="AAL21961.1"
/db_xref="GI:16421639"
/translation="MSDDMSGSPSSAGECGVLRSMQEVANSSQSEAKMLRTYINAMGNVYDREELHISVCPDPDPEARVADLAKVKAREQGRPLFCFPIOLHRLSSINAAKRRARESGYNGDFELVYPIKVNQHRVIESLIHSGPRLGLEGSAELMAVLAHAGMRYSVYNGCYKQDREYIRIALIGEMGHKVVLTIEKMEIATVEEARELNVPRICVARRIASOGSGKMOSSGGEKSKGLAQTQVLDLVERLPAGRLDSLOLHFLHLSQMANIDITATGYESARFVYVELHKLQGNVIOCFDGGGAGVDYEGTRSDCSVANGNEVANINIMATGDCERHGLPHPTVITSEGRVTAHHTVLYSNITVERNETDTPARAEVAPRALONLMTWEOEMHKPGTRRSRLVLRHDSQMDLHDHIGSSGAFSLOERMAQDPLELSCHEVOKOLDPONRAHRIIDELQERNADKKVYVNFSLFQSPAMGKIDOLFVPLRYLSDHEQVQAVLDITCDSGAIDHYIDGDIATMMPAEVDPEMPALGFPMFGYAVQELIGNNHNLFCGTEAVDVFVPPDGSVEVELSPEDQTVADMQLQVOLDPKTLTHFRQGVQTDLDLDAQOFLFEERAGLTYTLEDE"
complement(4467. .4472)
/gene="spea"
/note="putative RBS for spea: RegulonDB: STMS1H003127"
complement(4471. .4602)
/gene="yggB"
/note="STM3087"
complement(4471. .4602)
/gene="yggB"
/note="similar to E. coli ofr, hypothetical protein (AAC75976.1): Blastp hit to AAC75976.1 (48 aa), 73% identity in aa 6 - 46"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL21962.1"
/db_xref="GI:16421640"
/translation="MKKRPVQAQEHQRIQLLENPLVYGLLSRLRIAYVNCITLANKN"
complement(4531. .4539)
/gene="yggB"
/note="putative -10_signal for spea: RegulonDB: STMLTH004574"
complement(4550. .4558)
/gene="yggB"
/note="putative -35_signal for spea: RegulonDB: STMLTH004574"
4723. .4900
/gene="yggC"
/note="STM3088"
4723. .4728
/gene="yggC"

CDS	/note-"putative RBS for ygcC. RegulonDB: STMS1H003128"
	4736..4900
	/gene-"ygcC"
	/note-"similar to E. coli orf, hypothetical protein (AAC75977.1); Blastp hit to AAC75977.1 (71 aa), 78% identity in aa 1 - 51"
	/codon_start=1
	/transl_table=1
	/product-"putative cytoplasmic protein"
	/protein_id="AAL21965.1"
	/db_xref="GI:16421641"
	/translation-"MGITSAQMSQRDAGCGHARTRTMROIQONTVHYLVSPRRPRRKTINQAARF"
gene	complement(4897..5207)
	/gene-"ygd"
	/note-"STM3089"
	complement(4897..5196)
CDS	/gene-"ygd"
	/note-"similar to E. coli orf, hypothetical protein (AAC75978.1); Blastp hit to AAC75978.1 (83 aa), 52% identity in aa 1 - 78"
	/codon_start=1
	/transl_table=1
	/product-"putative inner membrane protein"
	/protein_id="AAL21964.1"
	/db_xref="GI:16421642"
	/translation-"MTEDSTSLKVDATGSLSVNOCYGINHTAILGHPFOAFRAFFGRNVGNQAVNSVDNADIMRIYALLFCVLGITRIDAKLLOKHTEPSSCRHFOPG"
RBS	complement(5202..5207)
	/gene-"ygd"
	/note-"putative RBS for ygd; RegulonDB: STMS1H003129"
	5237..6406
gene	/gene-"mek"
	/note-"STM3090"
RBS	5237..5242
	/gene-"mek"
	/note-"putative RBS for mek; RegulonDB: STMS1H003130"
CDS	5252..6406
	/gene-"mek"
	/EG_number="2.5.1.6"
	/note-"Adomet synthetase; similar to E. coli methionine adenosyltransferase 1 (Adomet synthetase); methyl and propylamine donor, corepressor of met genes (AAC75979.1); Blastp hit to AAC75979.1 (384 aa), 97% identity in aa 1 - 384"
	/codon_start=1
	/transl_table=1
	/product-"methionine adenosyltransferase 1"
	/protein_id="AAL21965.1"
	/db_xref="GI:16421643"
	W.
Query Match	80.5%; Score 929.4; DB 1; Length 22204;
Best Local Similarity	87.8%; Pred. No. 2e-212;
Matches 1014:	Conservative 0; Mismatches 141; Indels 0; Gaps 0
OY	1 ATGGCAAAACACCTTTTACGTCCGAGTCGCCGTCTGAAGGCGATCCTGACAAAATTGCT 60 Db 5252 ATGGCAAAACACCTTTTACGTCCGAGTCGCCGTCTGAAGGCGATCCTGACAAAATTGCT 60
OY	61 GACCAAATTCGATGCCTTTTAGCGCGGATCTCTGACACGATCCGAAAGCAGCGCTT 120 Db 5312 GACCAAATTCGATGCCTTTTAGCGCGGATCTCTGACACGATCCGAAAGCAGCGCTT 120
OY	121 GCCTGCGAAGCTACGTAATAAACCGGACGCGTTTTAGTTGGCGGCGAAATTCACACCAGC 180 Db 5372 GCCTGCGAAGCTACGTTAAACCGGACGCGTTTTAGTTGGCGGCGAATTCACACCAGC 180
OY	181 GCCGTGGTAGACATCGAAGAAGATCACCGCTTAACACCGCTTGCGAAATTTGGCTATGTGCAT 240 Db 5432 GCCGTGGTAGATCGAAGAAGATCACCGCTTAATACCGCTTGCGAAATTTGGCTATGTGCAT 240
OY	241 TCCGACATGGCTTGACGCTTAACGCTCTGCGCTTCAAGCGCATTCGGCAANAGTCT 300


```

Db 5492 TCCGACATGCGCTTGGACCACTCTTGGCGGCTACTGAGCGCAATTGGCAACAGTCT 5551
Oy 301 CCTGACATCAACAGGCGGCTTACCGTCCGATCCGCTGGAAAGCGGCGGTGACAG 360
Db 5552 CCGGATATCAACAGGCGGCTTACCGCGGACAGCCGCTGGAACAGCGCGGCGGACAG 5611
Oy 361 GCTGTGATCTTTGGCTACGCAACTATGAACGAGCGTCTGATGCGCACCTATACCC 420
Db 5612 GCGGTGATCTTTGGCTACGCAACTATGAACGAGCGTCTGATGCGCGGCAATACAGC 5671
Oy 421 TATGACACCGCTTGTGTACAGCGCTGAGTGCATTAAGGCAACGCGACTCGCTGG 480
Db 5672 TACGGGCAACCGTCTGTGTACAGCGTGAAGTGCATTAAGTGCACCGCTGCTATG 5731
Oy 481 CTGGCGCGGAGCGGCAAGGCGAGGCTTTTCAGTATGAGAGCGGCAAAATCGTGTG 540
Db 5732 TTGCGTCCGATGCAAAAAGCGAGCTCTTCAGTATGAGAGCGGCAAAATCGTGTG 5791
Oy 541 ATGATGCTGTGCTGCTTTCACCTGAGCACTGTGAAGAGATCGAGCAAAATCGCTCAA 600
Db 5792 ATGACGCGCGGTGCTTCTCTACGCAAGCAAGCAAGATGAGCAAAAATCGCTCAA 5851
Oy 601 GACGCGGTATGAGAGATCATCAAGCAATTCGCCGCTGATGCTGACTTGTGCC 660
Db 5852 GAAGCGGTGATGAGAGATCATCAAGCAATTCGCCGCTGATGATGATTAATAGTGC 5911
Oy 661 ACCAATTCCTGATCAACCGGCGTGTGCTTATGATGCGGCGGATGGGTGATGCTG 720
Db 5912 ACTAAGTCTTTATATCAACCAACCGGCGTGTGCTTATGATGCGGCGGATGGGTGCTG 5971
Oy 721 GCTGTGATGCTGTGCTTATATGCTGATGATGATGATGATGATGATGATGATGATG 780
Db 5972 GCTGTGATGCTGTGCTTATATGATGATGATGATGATGATGATGATGATGATGATG 6031
Oy 781 GGTGATGCTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 6032 GCGGCGCTTCTCGGTAAGATCCGCTAAAGTGAACCGCTTACGCTTACGCTGCGCT 6091
Oy 841 TATGTCGCAAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 6092 TATGTCGCAAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6151
Oy 901 TACGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 6152 TACGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6211
Oy 961 GTGCTTCTGCAACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 6212 GTGCTTCTGCAACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6271
Oy 1021 CTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 6272 TTTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6331
Oy 1081 TTTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 6332 TTTGATTCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6391
Oy 1141 GCGGCTCTGAAGTAA 1155
Db 6392 GCGGCTCTGAAGTAA 6406

```

```

RESULT 15
AL627277 230050 bp DNA linear BCT 06-JUN-2002
LOCUS AL627277
DEFINITION Salmonella enterica serovar Typhi (Salmonella strain CT18,
complete chromosome; segment 13/20.
ACCESSION AL627277 AL513382
VERSION AL627277.1 GI:16504049
KEYWORDS
SOURCE Salmonella enterica subsp. enterica serovar Typhi.
ORGANISM Salmonella enterica subsp. enterica serovar Typhi

```

```

REFERENCE
AUTHORS
1 (bases 1 to 230050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Comerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrah,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gara,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrall,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)
JOURNAL
MEDLINE
PUBMED
21534947
11677608
2 (bases 1 to 230050)
Parkhill,J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S\_typhi/).
FEATURES
source
location/Qualifiers
1..230050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
309..1598
/feature="STY3109"
/feature="sdac"
309..1598
/feature="STY3109"
/feature="sdac"
/feature="Fasta hit to Q46940 (409 aa), 40% identity in 416
aa overlap
Fasta hit to TDCC_ECOLI (443 aa), 50% identity in 442 aa
overlap
Orthologue of E. coli sdac (SDAC_ECOLI): Fasta hit to
SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
/codon_start=1
/transl_table=11
/product="putative serine transporter"
/protein_id="CAD02796.1"
/db_xref="GI:16504050"
/db_xref="SPTREMBL:O82435"
/transl_table=11
/translation="MKTTQSTFIASIDRSAMRKDTMMMLGLVGTAGVLPIN
ACVGMIPILIALIAPRTPFPAHGLTRFVLSGKNRCEIDTEVEEHPGAGKLT"
LYFPRIPIILVSVATINVESFLTHQALNPPRAILSLILVGMHTVIREOM
IVKANSILVFPVAVLMLDALYLIPONNGALLETSLFSDASVSTNGVLMWTLAIPV
VSEFNSPILISSFAVAKREYEGEAKCSKILARLIMVTVNFVFSVSLSTPA
DLAAKEDONISILSYLANAFNAPIIAMAIRKSFGLHIGARREGNGVIVK
SLRGGKSEINKLKITALFPLVLTVAATVATNPISILCMIEFLGPIIAMILFLPMY
ALOKPAMRKYSGHISNVFVIMGLAISALFYSLS"
1657..3024
/feature="STY3110"
/feature="sdab"
1657..3024
/feature="STY3110"
/feature="sdab"
/feature="Fasta hit to TDCC_ECOLI (454 aa), 73% identity in
454 aa overlap
Fasta hit to SDHL_ECOLI (454 aa), 77% identity in 455 aa
overlap
Orthologue of E. coli sdab (SDHL_ECOLI): Fasta hit to
SDHL_ECOLI (455 aa), 94% identity in 455 aa overlap"
/codon_start=1
/transl_table=11
/product="L-serine dehydratase 2 (L-serine deaminase 2)"
/protein_id="CAD02797.1"

```

gene
/db_xref="GI:16504051"
/db_xref="SPTREMBL:Q82434"
/translation="MISVFDFIKIGICGSSSHTVGPKKAGKQFTDDELIAHILITDVT
RVVDVYGLSLTGKCHNDLIALIMGLAGNLPDVTADIAIPFIQIDVTVHLLIAGNO
HVEFPPVCKSNFADNHLSTHENGMRITALAGNKLVSQRYISGGGFTIDEEHFGILT
NSEPANNVYPRKTTAADLDRHCQETGLSLGLMONEELALSHKLELBOHFPAKVEVNS
GIERVITTEGVLGCKLVRPRRAALRLRLVSQKTTDPMAVVDVMIMPLAVEENA
AGGRVATPPTAGCIPAVLAYDKFIREVNANSLARMLVVASIGSLYKMNASISG
AEVGCSAGCVASMAAGLAEILLGSPAOVCIAAEIAMEHNILITCDPAGVOVPC
IERNALISVKAIVNARMLRRTSPRVCIDKVIETMYETGDMNAKYRETSRGLAAK
ITVCD"
3135..3950
/gene="STY3111"
/note="exo"
3135..3950,"
/gene="STY3111"
/note="Orthologue of E. coli exo (EX9_ECOLI): Fasta hit to
EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
/codon_start=1
/transl_table=1
/product="5'-3' exonuclease"
/protein_id="CAD02798.1"
/db_xref="GI:16504052"
/db_xref="SPTREMBL:Q82433"
/translation="MLASPRECYPTRRFVRENTVAANLLIYDALNLIRIHAVGSP
CVETQALHDLIHSOPTHAVAVADDARSNGRHORLPYKGRPMPPDDLINERP
ALRAAFEGRGVCMASDGNEDDLAATLALVTGAQOATVSTDKYCOLLSFGRLI
RDYQKRLMDLPTIEKEGCVLPOLPDVGLAGISSKSPGVAGIGPSAKTOLLVQRO
NLGFIYALHDEYPEKWRKKLTNHEMAFLCHDIRLOTDLHIDNOQLRLVR"
3201..3947
/gene="STY3111"
/note="Pfam match to entry PF01367 5_3-exonuclease, 5'-3'
exonuclease, score 120.40, E-value 3.3e-32"
exonuclease(4042..5190)
/gene="STY3112"
/note="fuco"
complement(4042..5190)
/gene="STY3112"
/note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
372 aa overlap
Fasta hit to ADH2_ECOLI (383 aa), 44% identity in 371 aa
overlap
Orthologue of E. coli fuco (FUCC_ECOLI): Fasta hit to
FUCC_ECOLI (383 aa), 91% identity in 382 aa overlap"
/codon_start=1
/transl_table=1
/product="1,2-propanediol oxidoreductase (lactaldehyde
reductase)"
/protein_id="CAD02799.1"
/db_xref="GI:16504053"
/db_xref="SPTREMBL:Q82432"
/translation="MANRMILNENAFRGAVDALTDEYTRGCHKALITYDKLTQCC
GVVDVYTRMDAGLAMEITAGVIRPNPITISVYGQSLKFPYOSGADYLIALIGGSPDQT
KAVAGIISNNEFADVRLEGLSPTRKPSVPIYKALPTTAGTAACVITINVTDEKRR
KFCVCDPDIIDPAAFIIDAMDMGMPALKATGVADALTHALEGVTTPAAVALDIALI
KALEIADALGAVAGEKEACEMALGQYVAGKQFSVNGGLVGMNHPGAFINPFI
GVANAILPHVWRFNAGSTNKEFRDIAPAMCVKEGSLSEARNAAVEAFTLNRDVG
IPLHLDVGVKREDIPALDOAFDDVCTGMPRRASLADIYELNHTAM"
complement(4065..5169)
/gene="STY3112"
/note="Pfam match to entry PF00465 Fe-ADH, Iron-containing
alcohol dehydrogenases, score 655.70, E-value 2.4e-193"
alcohol dehydrogenases(4354..4416)
/gene="STY3112"
/note="PS00060 Iron-containing alcohol dehydrogenases
signature 2"
complement(4585..4671)
/gene="STY3112"
/note="PS00913 Iron-containing alcohol dehydrogenases
signature 1"
complement(5207..5854)
/gene="STY3113"
/note="fuca"

```

CDS
complement(5207..5854)
/gene="STY3113"
/note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in 179 aa overlap
Orthologue of E. coli fucA (FUC_A_ECOLI); Fasta hit to FUC_A_ECOLI (215 aa), 92% identity in 215 aa overlap"
/codon_start=1
/transl_table=1
/product="fuculose-1-phosphate aldolase"
/protein_id="CAD02800.1"
/db_xref="GI:16504054"
/db_xref="SPTREML:082431"
/translation="MEINRLRAQIIDTCLERLGLNQTAGNSVRYOQGLIIPTC
IPYELTSEHVFIDADQHEQGLPSEEMFPHMAVOTRPDAKVVNHNVHCAVAV
INPRLPIHYHMAACGNSIDPCAPATFEGRSLEDHVAALNKKAKTILDHQGLIAC
EENIDKALMLHAEVEVLAQLYSLTAIYDPPVUDEDEAIAIVLEFKRYGLRIE"
complement(5219..5845)
/gene="STY3113"
/note="Pfam match to entry PF00596 Aldolase-II, Class II
Aldolase and Adducin N-terminal domain., score 337.10,
E-value 2e-97"
6410..7726
/gene="STY3115"
/note="fucp"
6410..7726
/gene="STY3115"
/note="fucp"
6410..7726
/gene="STY3115"
/note="Orthologue of E. coli fucp (FUCP_ECOLI); Fasta hit
to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
/codon_start=1
/transl_table=1
/product="L-fucose permease"
/protein_id="CAD02801.1"
/db_xref="GI:16504055"
/db_xref="SPTREML:082430"
/translation="MGNTTITOTQSFRAVDAEQSKRRIIPALLCSLFPIIAVANNIN
NDILLPOQOAFITLTFNAGLIQSAFVFGYVPIPAAILKKLSYKGIITGLVLY
GALFLPFAEAIIMNTYLFILGLFIIAAGLEETANFPVVALGESGHRRLNLAOT
FNSGAIIAVVGOSLISNVHQSQELDKMTPOLSAVYKHSILSVOTPYMIIVA
VLVAAILMLTKPALQSDHSDKQSSFLSSRLIIRHMRVAVLOFCVGOATQ
CWSLIRAIYAEIEMCTGTPGFANYLTGTMNCFTICRPGMLISFEAHKILAAVAL
AMLICLISAFESGHIGLALTLTCLSPKSNIOYPTFSLCINKLDGOTKGGSFIVMTII
GGGIVTPMGVFSVDAAGIIPYAEIVLPALFPIVIFLAFRISOAATN"
7758..9488
/gene="STY3116"
/note="fucI"
7758..9488
CDS
Query Match 80.3% Score 927.8 DB 1: Length 230050:
Best Local Similarity 87.7% Pred. No. 5,3e-212:
Matches 1013; Conservative 0; Mismatches 142; Indels 0; Gaps 0:
1 ATGGCAAAACACCTTTTACGTCCGACGRCCTCTCGAAGGCACTTCACAAAATTCGT 60
|||||
DB 131377 ATGGCAAAACACCTTTTACGTCTAGTCTCGATACAGAAGGCACTTCACAAAATTCGCA 131346
61 GACCAAAATTTCTGATGCCCTTTTAGACGGCATCTCGACAGAGATCCGAACACCGCTT 120
|||||
DB 131437 GACCAAAATTTCTGATGCCCTTTTAGACGGCATCTCGACAGAGATCCGAACACCGCTTC 131456
121 GCTTCGAAACCTACGTAAAAACCGGCATGTTTTAGTTGGGGCGCAATACACACACG 180
|||||
DB 131497 GCTTCGAAACCTACGTAAACCGGCATGTTTTAGTTGGGGCGGTACATCACACACG 131556
181 GCGTGGGTAGACATCGAAGACATCACCGGTACACACCGCTTCGCAAAATTTGGCTATGTGCAT 240
DB 131557 GCGTGGGTAGATTCGAAGACATTCACCGGTATACGGTCCGCAAAATTTGGCTATGTGCAT 131616
241 TCCGACATGGGCTTTCAGCGCTTAACCTCTCTGGGGTTCTAGAGCGTATCGCAACACGCT 300
DB 131617 TCCGACATGGGCTTTCAGCGCTTTCAGCGCTTCTGGGGTTCTAGAGCGTATCGCAACACGCT 131676
301 CCTGACATCAACACGAGCGCTTACCGCTGCTGATCCGCTGGAAACAGGCGCGGGTGACAG 360

```

Dp	1316177	CCGGATATATCAACGACGGGGCGTTGACCCGGGACAGACCCCGTGGAAACAAGGGGGGCGACACAG	131736
Oy	361	GGTGTGATGTTTGGCTACCGCACTAATGAACCGGAGCTGCTATGCCAGACCTATACAC	420
Dp	131737	GGCGTGTATGTTTGGCTCTCGGGACACAAATGAAGAACCGAGCTACTGTATGGCTGGCCAAATCC	131796
Oy	421	TATGCACACCGCTGTGTACAGACGCTCAGAGGTGAAGTGGCTAAAGAGGCACTCTGGCCGTGG	480
Dp	131797	TACCGGCAACCGTCTGTATACACCGCTCAGCGCTGAAGTGGCTAAATATGGCAACCTCCGATGG	131856
Oy	481	CTGGCCCCGGACGGCAAAAGCCAGGTGACTTTTCATGTATGACGACGGCAAAATCGTTGGT	540
Dp	131857	TTGGCTCCGGATGCAAAAAGCCAGAGTCACTTCCAGATATGACAGCGCAAAATCGTCGGT	131916
Oy	541	ATGCATCTGTGCGGCTTTCCACTCAGCACTGTGAAGATGCAACCAAAATCGCTGCAAA	600
Dp	131917	ATGCACCCCGGTGTTCTCTCTACCGCACCGCAACGAAATATGCAACCAAAATGCTGCAAA	131976
Oy	601	GAACGGGTAAATGAAGAGATCATCAACCAATTCGGCCGCTGAATGGCTGACTTCTGCC	660
Dp	131977	GAACGGGTGAATGAAGAGATCATCAAAACCAATTCGTGGCGTGTGAATGGTTAAATACGTGC	132036
Oy	661	ACCAAAATTTCTTATCAACCCGACCGGTGCTTTGGTTATCGGTGGCCCAATGGGTACTGC	720
Dp	132037	ACTAAGTTTATATCAACCCAGCCAGGCGCTTTGTTATTCGGCGCCCGAATGGGCGATTGG	132096
Oy	721	GGTCTGACTGGTCTCTAAATATTATCGTTGATACGTACGGCGGCAATGGCGGCTCACGGTGGC	780
Dp	132097	GGTCTGACCGGTGCTTAAGATCATCTGTGATACTACGGCGGTATGGCGGCTCACGGCGGC	132156
Oy	781	GGTGCATTTCTTGTTAAAGATCCATCAAAAGTGGACCGTTCCGACGCTACGCGAGCACT	840
Dp	132157	GGCGCTTTCTTCGCTAAAGATCCGCTCTAAAGTTGACCGCTTCTCAGCTTTCGCTGGCGGT	132216
Oy	841	TATGTCCGGAAACATCGTTGGTGGCTGGGCTGGCGCATGCTGTGTAATTTCAAGTTCG	900
Dp	132217	TATGTCCGGAAACATCGTGGGCGGACGAGTCTGGCGATGCTGTGTAATTCAGGTTCTC	132276
Oy	901	TACGCATTCGGCGTGGCTGGAACCGACCTCCATCATGGTATAGAAATTTTCGCTACTGGA	960
Dp	132277	TACGCTATCGCGCTGACGGAGACCGACGTCATCATGGTATGGAACGTTCCGCTACTGAAAA	132336
Oy	961	CTGCGCTTCTGAACAATGACCTCTGCTGTGATGTAAGTTCTTTCGACCTCGGCCCATACGGT	1020
Dp	132337	GTCGCTGTGTGAACAGTTGATTTCTGCTGTGCTGGCGCAATTTCTTGACACCTCGGCCGTGGG	132396
Oy	1021	CTGATTCAGATGCTGATCTGCTGCACCGCATCTCAAAAGAACCGGACGATACGGCTAC	1080
Dp	132397	TTGATTCAGATGCTGATCTGCTCTACCCGATCTACAAAGAAACCTGGCGCTTACGGCCAC	132456
Oy	1081	TTTGGTCTGTGAACATTTTCCGTGGGAAAAAACCGACAAGCCGACGCTGCTGCCGATGCT	1140
Dp	132457	TTTCGGTCCGAAACCTTCCATGGGAAAAAACCGACAAGCCGACGATGCTGCCGATGCT	132516
Oy	1141	GCCGCTGTGAAGTAA 1155	
Dp	132517	GCCGCTGTGAAGTAA 132531	

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 08:16:32 : Search time 290.095 Seconds
(without alignments)
8966.219 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaacacccctttac.....atgcgtccgctcgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.GeneSeq_101002:*

- 1: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1155	100.0	1155	20	AA209784	E. coli S-adenosyl
2	1155	100.0	1155	22	AAAS9433	Met K oligonucleot
3	1155	100.0	1155	23	AAAS52561	E. coli DNA for ce
4	1155	100.0	3720	20	AA209790	Plasmid pHS1 meck
5	1155	100.0	4975	20	AA209792	Plasmid pHS1 meck/
6	1155	100.0	12354	22	AAAS4243	DNA encoding novel
7	1155	100.0	12354	22	AAAS4261	DNA encoding novel
8	1155	100.0	12354	22	AAAS4267	DNA encoding novel
9	926.8	80.2	1167	23	AAAS56004	Salmonella typhi D

10	815.4	70.6	2853	23	AAAS6725	DNA encoding novel
11	641.4	55.5	1155	23	AAAS3462	Haemophilus influe
12	586.2	50.8	1191	23	AAAS54057	Pseudomonas aerugi
13	533.4	46.2	2577	23	AAAS90256	DNA encoding novel
14	533.4	46.2	2577	23	AAAS94403	DNA encoding novel
15	505.8	43.8	3219	21	AAAS20961	E. coli encoding novel
16	503.4	43.6	56485	21	AAAS1476	N. meningitidis bet
17	503.4	43.6	349980	21	AAAS1612	N. meningitidis pa
18	503.4	43.6	837096	21	AAAS1489	N. meningitidis pa
19	463.4	40.1	1149	22	ABAS9246	Escherichia coli p
20	463.4	40.1	27324	22	ABAS9246	Escherichia coli p
21	460	39.8	62909	22	AAAS23545	Genomic fragment
22	391.4	33.9	1203	24	ABK71681	Bacillus lichenifo
23	375.4	32.5	1191	23	AAAS1934	Staphylococcus aur
24	375.4	32.5	1296	23	AAAS1456	Staphylococcus aur
25	375.4	32.5	1296	23	AAAS5121	Staphylococcus aur
26	374	32.4	4848	24	AAAS2684	Streptomyces fradi
27	374	32.4	4848	24	AAAS2684	Streptomyces fradi
28	371	32.1	2365589	24	ABAS0521	Genomic sequence
29	370.6	32.1	1185	23	AAAS1032	Enterococcus faeca
30	369.8	32.0	1191	23	AAAS5678	Streptococcus pneu
31	369.8	32.0	5392	15	AAAS64201	Sequence comp:isin
32	368.2	32.0	1158	23	AAAS5241	Enterococcus faeca
33	368.2	31.9	1191	21	AAAS1552	Streptococcus pneu
34	368.2	31.9	1191	23	AAAS5856	Streptococcus pneu
35	368.2	31.9	1208	15	AAAS4204	snac gene encoding
36	368.2	31.9	8494	19	AAAS5296	Streptococcus pneu
37	361.8	31.3	10320	18	AAAS74454	Streptococcus pneu
38	356.6	30.9	1377	20	AAAS1609	Enterococcus faeca
39	355.4	30.8	1194	24	ABAS6257	Streptococcus pneu
40	351.4	30.4	31702	24	ABAS6190	Listeria innocua c
41	349.2	30.2	1248	24	ABAS7777	Streptococcus epi
42	347.6	30.1	1210	22	AAAS4976	S. epidermidis gen
43	333.6	28.9	16710	20	AAAS20544	Polynucleotide seq
44	332.6	28.8	1194	24	ABAS68256	Streptococcus poly
45	332.2	28.8	1283	22	AAAS91225	Human methionine a

ALIGNMENTS

RESULT 1	AA209784	standard: DNA: 1155 BP.
ID	AA209784	
AC	AA209784:	
XX		
DT	23-NOV-1999	(first entry)
DE	E. coli S-adenosyl methionine synthase SAM DNA	
XX		
KW	SAM: S-adenosyl methionine synthase; biosynthesis; biotin; bios1; bios2;	
KW	bios3; cofactor; decarboxylation; vitamin H; ss.	
OS	Escherichia coli.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1155
FT		/*tag= a
FT		/product= "S-adenosyl methionine synthase"
PN	DEL19806872-A1.	
PD	26-AUG-1999.	
XX		
PF	19-FEB-1998:	98DE-1006872.
XX		
XX	19-FEB-1998:	98DE-1006872.
XX		
PA	(BADI) BASF AG.	
XX		
PI	Schroeder H:	
XX		

DNA encoding novel
Haemophilus influe
Pseudomonas aerugi
DNA encoding novel
DNA encoding novel
E. coli encoding novel
N. meningitidis bet
N. meningitidis pa
N. meningitidis pa
Escherichia coli p
Escherichia coli p
Genomic fragment
Bacillus lichenifo
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Streptomyces fradi
Streptomyces fradi
Genomic sequence
Enterococcus faeca
Streptococcus pneu
Sequence comp:isin
Enterococcus faeca
Streptococcus pneu
Streptococcus pneu
snac gene encoding
Streptococcus pneu
Streptococcus pneu
Enterococcus faeca
Streptococcus pneu
Listeria innocua c
Streptococcus epi
S. epidermidis gen
Polynucleotide seq
Streptococcus poly
Human methionine a

DR WPI: 1999-480095/41.
 DR P-PSDB: AAY33263.
 PT Production of biotin by expressing S-adenosyl-methionine synthase and
 PT second biotin synthase gene in host cells
 XX
 PS Claim 1: Page 8-10; 48pp; German.
 XX
 CC This invention describes a novel method for the preparation of biotin
 CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
 CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
 CC sequence (1), and (b) at least one of the other biotin biosynthesis
 CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
 CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
 CC most animals and some microorganisms. Expression of biotin plus bios1,
 CC bios2 or bios3 significantly increases productivity of biotin
 CC biosynthesis, particularly by at least 3 times. This sequence encodes
 CC the Escherichia coli SAM protein which is used in the method of the
 CC invention.
 CC
 XX
 SO Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other:
 Query Match 100.0%; Score 1155; DB 20; Length 1155;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 0;
 Oy 1 ATGGCAAAACACCTTTTACGTCGCGAGTCGCTCTGTAAGGCGATCTGACAAATTTGCT 60
 Db 1 ATGGCAAAACACCTTTTACGTCGCGAGTCGCTCTGTAAGGCGATCTGACAAATTTGCT 60
 Oy 61 GACCAATTTCTGATGCGCGTTTAAAGCGGATCCTCGAAGAGATCCGGAAGCAGCGGTT 120
 Db 61 GACCAATTTCTGATGCGCGTTTAAAGCGGATCCTCGAAGAGATCCGGAAGCAGCGGTT 120
 Oy 121 GCTTGGCAAAACCTTACGTCGCGAGTCGCTCTGTAAGGCGATCTGACAAATTTGCT 180
 Db 121 GCTTGGCAAAACCTTACGTCGCGAGTCGCTCTGTAAGGCGATCTGACAAATTTGCT 180
 Oy 181 GCTTGGCGTAGACATGGAAGATACCGCTTACACCGCTTGGCAATTTGCTATGTCAT 240
 Db 181 GCTTGGCGTAGACATGGAAGATACCGCTTACACCGCTTGGCAATTTGCTATGTCAT 240
 Oy 241 TCCGACATGGGCTTGAAGCGTAACTCTGTCGCGGTTCTGAGCGGTATTCGGCAACAGTCT 300
 Db 241 TCCGACATGGGCTTGAAGCGTAACTCTGTCGCGGTTCTGAGCGGTATTCGGCAACAGTCT 300
 Oy 301 CCTGACATGCAACGAGGCGTTGACCGTTCGCGATCCGCTGGAACAGGCGGCGTACAC 360
 Db 301 CCTGACATGCAACGAGGCGTTGACCGTTCGCGATCCGCTGGAACAGGCGGCGTACAC 360
 Oy 361 GCTCTGATGTTTGGCTAGCGCACTAATGAACCGACGTCGTGATGCGACGACTATCACC 420
 Db 361 GCTCTGATGTTTGGCTAGCGCACTAATGAACCGACGTCGTGATGCGACGACTATCACC 420
 Oy 421 TATGACACCGCTCTGTAACGCTCAGGCTCAAGTGGCTAAACCGGACACTCTCCCTGG 480
 Db 421 TATGACACCGCTCTGTAACGCTCAGGCTCAAGTGGCTAAACCGGACACTCTCCCTGG 480
 Oy 481 CTGCGCCGAGCGGAAAGCGAGTACTTTTCAATGATGACGAGCGCAAAATCGTTGGT 540
 Db 481 CTGCGCCGAGCGGAAAGCGAGTACTTTTCAATGATGACGAGCGCAAAATCGTTGGT 540
 Oy 541 ATCGATGCTGTCTCTTCCACTCAGCAGCTCTGAAGAGATGACGAGAAATTCCTGCAA 600
 Db 541 ATCGATGCTGTCTCTTCCACTCAGCAGCTCTGAAGAGATGACGAGAAATTCCTGCAA 600
 Oy 601 GAACGGTAATGGAAGAGATATCAAGCAATTTCTGCCCGCTGAATGGCTGACTTTCGCC 660
 Db 601 GAACGGTAATGGAAGAGATATCAAGCAATTTCTGCCCGCTGAATGGCTGACTTTCGCC 660
 Oy 661 ACCAAATTTCTTCAATCAACCGGAGCGTCTTCTTATACGTGAGCGCAATGGTGACTGC 720
 Db 661 ACCAAATTTCTTCAATCAACCGGAGCGTCTTCTTATACGTGAGCGCAATGGTGACTGC 720

Oy 721 GGTCTGACTGCTGCTGTAATTTATCTGTTGATACCTACGCGGCAATGCGGCTACGCTGCC 780
 Db 721 GGTCTGACTGCTGCTGTAATTTATCTGTTGATACCTACGCGGCAATGCGGCTACGCTGCC 780
 Oy 781 GGTGCAATCTCTGCTGTAAGATTCATCAAAAGTGAGACGTTCCGACGCTTACGAGCAGT 840
 Db 781 GGTGCAATCTCTGCTGTAAGATTCATCAAAAGTGAGACGTTCCGACGCTTACGAGCAGT 840
 Oy 841 TATGCGCGAATAACATCTGCTGCTGCTGCGCTGCGGATGCTTGTGAATTCAGTTTCC 900
 Db 841 TATGCGCGAATAACATCTGCTGCTGCTGCGCTGCGGATGCTTGTGAATTCAGTTTCC 900
 Oy 901 TACGCAATCGCGCTGCTGCTGACCGACCTCATATGCTAGTAAGAACTTTCGCTACTGAGAA 960
 Db 901 TACGCAATCGCGCTGCTGCTGACCGACCTCATATGCTAGTAAGAACTTTCGCTACTGAGAA 960
 Oy 961 GTGCGTCTGTAACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 961 GTGCGTCTGTAACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Oy 1021 CTGATTCAGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 Db 1021 CTGATTCAGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 Oy 1081 TTTGCTGCTGTAACATTTCCGCTGCGGAAAGAACGACAAAGCGTCTGCTGCTGCTGCT 1140
 Db 1081 TTTGCTGCTGTAACATTTCCGCTGCGGAAAGAACGACAAAGCGTCTGCTGCTGCTGCT 1140
 Oy 1141 GCGGCTGTAAGTAA 1155
 Db 1141 GCGGCTGTAAGTAA 1155
 RESULT 2
 AAA99433
 ID AAA99433 standard; DNA: 1155 BP.
 AC AAA99433:
 AC 26-JAN-2001 (first entry)
 DT
 DE Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
 XX
 KM L-methionine production; Met; Escherichia coli; ds.
 XX
 OS Escherichia coli.
 OS
 PN JP2000139471-A.
 XX
 PD 23-MAY-2000.
 PD
 PF 17-NOV-1998: 98JP-0326717.
 PF
 PR 17-NOV-1998: 98JP-0326717.
 PR
 PA (AJIN) AJINOMOTO KK;
 PA
 DR WPI: 2001-018703/03.
 DR P-PSDB: AAB26814.
 PT Fermentative process for preparation of L-methionine (Met), comprises
 PT using a modified Met producing microorganism, particularly an
 PT Escherichia species
 XX
 PS Example 2: Page 17-18; 23pp; Japanese.
 CC The invention relates to a process for preparing L-methionine (Met)
 CC using a modified Met producing organism. The modified microorganism has
 CC its Met biosynthetic system repressor deleted, and has enhanced
 CC homoserine transuccinylase activity. The organism optionally has an
 CC attenuated intracellular S-adenosyl methionine synthetase (SAM). The
 CC process is used for the preparation of L-methionine. Sequences AAA99416

CC and AAB9433 encode proteins represented in AAB26816 and AAB26814
 CC respectively, which are involved in the Escherichia coli Met biosynthetic
 CC pathway. AAB9433 (excluding AAB9433) represent primers. All
 CC sequences are used in the production of the modified Met producing
 CC organism of the invention.

XX Sequence 1155 BP: 277 A: 312 C: 303 G: 263 T: 0 other:

Query Match 100.0%: Score 1155: DB 22: Length 1155:
 Best Local Similarity 100.0%: Pred. No. 0:
 Matches 1155: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 1 ATGCAAAACACCTTTTACGTCAGTCCTCTGAGAGGCGATCTGACAAAATTGCT 60
 Db 1 ATGCAAAACACCTTTTACGTCAGTCCTCTGAGAGGCGATCTGACAAAATTGCT 60
 Oy 61 GACCAAAATTTCTGATGCCGTTTATAGCGCATCTCGAAGAGATCCGAACACGCGTT 120
 Db 61 GACCAAAATTTCTGATGCCGTTTATAGCGCATCTCGAAGAGATCCGAACACGCGTT 120
 Oy 121 GCTTCCGAAACCTAGCTAAAAACCGGCGATGTTAGTTGGCGGGAATCCACACGAC 180
 Db 121 GCTTCCGAAACCTAGCTAAAAACCGGCGATGTTAGTTGGCGGGAATCCACACGAC 180
 Oy 181 GCGTGGTAGACATCGAAGAGATCACCGCTTAACACCGCTTCCGGAATTTGCTATGTCAT 240
 Db 181 GCGTGGTAGACATCGAAGAGATCACCGCTTAACACCGCTTCCGGAATTTGCTATGTCAT 240
 Oy 241 TCCGACATGGGCTTTGACGCTAATCTCTGCGGCTTCTGAGCGCTATCGGCAACAGTCT 300
 Db 241 TCCGACATGGGCTTTGACGCTAATCTCTGCGGCTTCTGAGCGCTATCGGCAACAGTCT 300
 Oy 301 CCTGACATCAACGAGGGGTTGACCGTCCGATCGGCTGGAACAGGCGGCGTACACGAG 360
 Db 301 CCTGACATCAACGAGGGGTTGACCGTCCGATCGGCTGGAACAGGCGGCGTACACGAG 360
 Oy 361 GGTCTGATGTTGGCTTACGCAACTAATGAAACCGAGCTGATGCCAGACATCTACAC 420
 Db 361 GGTCTGATGTTGGCTTACGCAACTAATGAAACCGAGCTGATGCCAGACATCTACAC 420
 Oy 421 TATGACACACCGCTGCTGACAGCGTGAAGTCCGTAATAACGCGACTTCCGCTGG 480
 Db 421 TATGACACACCGCTGCTGACAGCGTGAAGTCCGTAATAACGCGACTTCCGCTGG 480
 Oy 481 CTGCGCCCGGAGCGCAAAAGCCAGTCTTTCAGTATGACGAGGCAAAATTCGTTGCT 540
 Db 481 CTGCGCCCGGAGCGCAAAAGCCAGTCTTTCAGTATGACGAGGCAAAATTCGTTGCT 540
 Oy 541 ATCGATGCTGCTGCTTTCCACTCAGCAGCTCTGAAGAGATCGACAGAAATCGCTGCA 600
 Db 541 ATCGATGCTGCTGCTTTCCACTCAGCAGCTCTGAAGAGATCGACAGAAATCGCTGCA 600
 Oy 601 GAGCGGTAATGAAGAGATCATCAAGCCATTTGCGCCGCTGAATGGCTGACTTGGC 660
 Db 601 GAGCGGTAATGAAGAGATCATCAAGCCATTTGCGCCGCTGAATGGCTGACTTGGC 660
 Oy 661 ACCAAATTTCTATCAACCCGAGCGGCTGTTTCTGTTATCGGAGCCCAATGGGTACTGC 720
 Db 661 ACCAAATTTCTATCAACCCGAGCGGCTGTTTCTGTTATCGGAGCCCAATGGGTACTGC 720
 Oy 721 GCTCTGACGCTGCTAAATTTATCGTTGATACCTACGCGGCGATGCGCTGACGCTGC 780
 Db 721 GCTCTGACGCTGCTAAATTTATCGTTGATACCTACGCGGCGATGCGCTGACGCTGC 780
 Oy 781 GGTGATTTCTCTGTAAGAGATCATCAAAAGTGAACCGTTCGCGAGCTTACGACACGCT 840
 Db 781 GGTGATTTCTCTGTAAGAGATCATCAAAAGTGAACCGTTCGCGAGCTTACGACACGCT 840
 Oy 841 TATGTCGGGAAACATGTTGCTGCTGCGCGGCGATCGTTGGAATTCAGGTTTCC 900
 Db 841 TATGTCGGGAAACATGTTGCTGCTGCGCGGCGATCGTTGGAATTCAGGTTTCC 900
 Oy 901 TACGCAATCGCGCTGCTGAACCGACCTTCATCATGTGTAAGAACTTTCGTTACTGAGAA 960

Db 901 TACGCAATCGCGCTGCTGAACCGGCTGCAATCATGTGTAAGAACTTTCGTTACTGAGAA 960
 Oy 961 GTGCTTTCTGAACACCTACCTGCTGAGCTGATGATTTCTTGACACTCGGCGCATACGCT 1020
 Db 961 GTGCTTTCTGAACACCTACCTGCTGAGCTGATGATTTCTTGACACTCGGCGCATACGCT 1020
 Oy 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCTCAAGAGAACCGGACGATACGCTGAC 1080
 Db 1021 CTGATTCAGATGCTGATCTGCTGACCGCATCTCAAGAGAACCGGACGATACGCTGAC 1080
 Oy 1081 TTTGCTGCTGAACATTTCCCGTGGGAAACCGACGAAAGGCTGCTGCGGATGCT 1140
 Db 1081 TTTGCTGCTGAACATTTCCCGTGGGAAACCGACGAAAGGCTGCTGCGGATGCT 1140
 Oy 1141 GCGGCTCTGAAGTAA 1155
 Db 1141 GCGGCTCTGAAGTAA 1155

RESULT 3

AAS52561
 ID AAS52561 standard; DNA: 1155 BP.

XX AAS52561:

XX 13-FEB-2002 (first entry)

XX E. coli DNA for cellular proliferation protein #283.

XX Antisense: ds: prokaryotic cellular proliferation gene;
 XX antibiotic: antibacterial; drug design.

XX Escherichia coli.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JN, Wall D, Traxlick JD, Carr CJ:
 PI Yamamoto RT, Xu HH:
 XX WPI: 2001-611495/770.
 DR P-PSDB: AAU34702.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Claim 27: Seq ID No 6198; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, pseudomonas aeruginosa and enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1155 BP: 277 A: 312 C: 303 G: 263 T: 0 other:
 Query Match 100.0%: Score 1155: DB 23: Length 1155:
 Best Local Similarity 100.0%: Pred. No. 0:
 Matches 1155: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 ATGGCAAAACACCTTTTACGTCGAGTCCGCTCTGTAAGGCGCTGACAAATTGCT 60
 DB 1 ATGGCAAAACACCTTTTACGTCGAGTCCGCTCTGTAAGGCGCTGACAAATTGCT 60
 OY 61 GACCAATTTCTGATGCCGTTTACGCGCATCTCGAAGAGATCCGAAAGCAGCGGCT 120
 DB 61 GACCAATTTCTGATGCCGTTTACGCGCATCTCGAAGAGATCCGAAAGCAGCGGCT 120
 OY 121 GCTTGGCAAAACCTGTAAGAAACCGGATGTTTATGTTGGCGGCAAAATCACCAGC 180
 DB 121 GCTTGGCAAAACCTGTAAGAAACCGGATGTTTATGTTGGCGGCAAAATCACCAGC 180
 OY 181 GCCTGGGTAGACATCGAAGAGATCACCCTGACACCGTTGCGAAATTGGCTATGTCAT 240
 DB 181 GCCTGGGTAGACATCGAAGAGATCACCCTGACACCGTTGCGAAATTGGCTATGTCAT 240
 OY 241 TCCGACATGGGCTTTGACGCGTAACCTCGTCCGCTGTCGAGCGCTATCGGCAACACT 300
 DB 241 TCCGACATGGGCTTTGACGCGTAACCTCGTCCGCTGTCGAGCGCTATCGGCAACACT 300
 OY 301 CCTGACATCAACAGGCGCTGACGCGTACCGCTGTAAGCGGCGCGGTGACAC 360
 DB 301 CCTGACATCAACAGGCGCTGACGCGTACCGCTGTAAGCGGCGCGGTGACAC 360
 OY 361 GGTCTGATGTTGGCTACGCACTAATGAACCGACGCTGATGCCAGCACTATAC 420
 DB 361 GGTCTGATGTTGGCTACGCACTAATGAACCGACGCTGATGCCAGCACTATAC 420
 OY 421 TATGCAACGCGTGTGATGACGCGTACGCGTGAAGTGGTAAACCGGCGACTCCGCG 480
 DB 421 TATGCAACGCGTGTGATGACGCGTACGCGTGAAGTGGTAAACCGGCGACTCCGCG 480
 OY 481 CTGGCGCGGACGCGAAAGCGAGTGAATTTTCAATATGACGCGCAAAATCGTTGGT 540
 DB 481 CTGGCGCGGACGCGAAAGCGAGTGAATTTTCAATATGACGCGCAAAATCGTTGGT 540
 OY 541 ATGCAATGCTGTGCTGCTTTCACCTCAGCACTCTGAAGAGATGACAGAAATCGCTG 600
 DB 541 ATGCAATGCTGTGCTGCTTTCACCTCAGCACTCTGAAGAGATGACAGAAATCGCTG 600
 OY 601 GAAGCGGTAATGAAGAGATCATCAAGCAATTTGCGCGGCTGAATGGCTACTTTC 660
 DB 601 GAAGCGGTAATGAAGAGATCATCAAGCAATTTGCGCGGCTGAATGGCTACTTTC 660
 OY 661 ACCAAATCTTCATCAACCGCAGCGGCTGTTGCTTATCGTGGCCCAATGGTGACTGC 720
 DB 661 ACCAAATCTTCATCAACCGCAGCGGCTGTTGCTTATCGTGGCCCAATGGTGACTGC 720
 OY 721 GGTCTGACTGCTGTAATTAATGTTGATACCTACGCGGCGCATGGCGCTCAGCGTGGC 780
 DB 721 GGTCTGACTGCTGTAATTAATGTTGATACCTACGCGGCGCATGGCGCTCAGCGTGGC 780
 OY 781 GGTGATCTCTGTAAGAGATCATCAAAAGTGAACCTTCCGACGCTTACGACGACGT 840
 DB 781 GGTGATCTCTGTAAGAGATCATCAAAAGTGAACCTTCCGACGCTTACGACGACGT 840
 OY 841 TATGCGGCAAAACATGTTGCTGCTGCGCGCATCTGTTGGAATTCAGATTGCC 900
 DB 841 TATGCGGCAAAACATGTTGCTGCTGCGCGCATCTGTTGGAATTCAGATTGCC 900

DB 841 TATGCGGCAAAACATGTTGCTGCTGCGCGCATCTGTTGGAATTCAGATTGCC 900
 OY 901 TACCAATGCGCGCTGCTGTAACCGACCTCCATATGTAAGAACTTTCGCTACTGAGAA 960
 DB 901 TACCAATGCGCGCTGCTGTAACCGACCTCCATATGTAAGAACTTTCGCTACTGAGAA 960
 OY 961 GTGCGCTTCTGAACACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 961 GTGCGCTTCTGAACACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 OY 1021 CTGATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 1021 CTGATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 OY 1081 TTTGGCTGTAACATTTTCCGCTGGGAAAAACCGACAAACCGACGCTGCTGCCGATGCT 1140
 DB 1081 TTTGGCTGTAACATTTTCCGCTGGGAAAAACCGACAAACCGACGCTGCTGCCGATGCT 1140
 OY 1141 GCGCGTCTGAAGTAA 1155
 DB 1141 GCGCGTCTGAAGTAA 1155

RESULT 4
 AA209790
 ID AA209790 standard; DNA: 3720 BP.
 XX
 AC AA209790:
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Plasmid pHS1 metk DNA.
 XX
 KW SAM: S-adenosyl methionine synthase: biosynthesis; biotin; bios2;
 KW bios3; cofactor; decarboxylation; Vitamin H; metk; ss.
 XX
 OS Synthetic.
 FH Key Location/Qualifiers
 FT 530..1684 /*tag= a
 FT COS /*product= "metk"
 XX
 PN DEL9806872-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 19-FEB-1998; 98DE-1006872.
 XX
 PR 19-FEB-1998; 98DE-1006872.
 XX
 PA (BADI) BASF AG.
 XX
 PI Schroeder H:
 XX
 DR WPI: 1998-480095/41.
 DR P-PSDB: AAT33267.
 XX
 PT Production of biotin by expressing S-adenosyl-methionine synthase and
 PT second biotin synthesis gene in host cells -
 XX
 PS Example 1c; Page 25-26; 48bp; German.
 XX
 CC This invention describes a novel method for the preparation of biotin
 CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
 CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
 CC sequence (1), and (b) at least one of the other biotin biosynthesis
 CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
 CC (de)carboxylation reactions and is an essential Vitamin (Vitamin H) for
 CC cell animals and some microorganisms. Expression of biotin plus bios1,
 CC bios2 or bios3 significantly increases productivity of biotin
 CC biosynthesis, particularly by at least 3 times. This sequence encodes
 CC the metk protein found in plasmid pHS1 which is used in the method of

CC the invention.
 XX
 SO Sequence 3720 BP: 932 A: 1019 C: 954 G: 815 T: 0 other:
 Query Match 100.0%; Score 1155; DB 20; Length 3720;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1155: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGCAAAACACCTTTTTCGTCCGAGTCCGCTCTGAGGGGATCCACAAAATTGCT 60
 DB 530 ATGGCAAAACACCTTTTTCGTCCGAGTCCGCTCTGAGGGGATCCACAAAATTGCT 589
 OY 61 GACCAAAATTTCTGATGCCCTTTTATAGACGGATCTCGAAGAGATCCGAACAGCGCTT 120
 DB 590 GACCAAAATTTCTGATGCCCTTTTATAGACGGATCTCGAAGAGATCCGAACAGCGCTT 649
 OY 121 GCTTGGCAACCTACGTAAACACGGCATGTTTATGTTGCGCGCAAAATCACACACAGC 180
 DB 650 GCTTGGCAACCTACGTAAACACGGCATGTTTATGTTGCGCGCAAAATCACACACAGC 709
 OY 181 GGCTGGGTAGACATGCAAGATCAACCCCTAACACCCCTTCGGGAATTTGGCTATGTGAT 240
 DB 710 GGCTGGGTAGACATGCAAGATCAACCCCTAACACCCCTTCGGGAATTTGGCTATGTGAT 769
 OY 241 TCGACATGAGCGTTTGACGCTAACTGCTGCGGTTCTGAGCGCTATCGCAAAACAGTCT 300
 DB 770 TCGACATGAGCGTTTGACGCTAACTGCTGCGGTTCTGAGCGCTATCGCAAAACAGTCT 829
 OY 301 CCTGACATCAACCGAGCGCTTGACCGCTGCCGATCCGCTGCAACAGCGCGCGCTGACAG 360
 DB 830 CCTGACATCAACCGAGCGCTTGACCGCTGCCGATCCGCTGCAACAGCGCGCGCTGACAG 889
 OY 361 GGCTGATGTTGGGTGCTAGCACTAATGAAACCGAGCGTCTATGCGCAGCCTATACAC 420
 DB 890 GGCTGATGTTGGGTGCTAGCACTAATGAAACCGAGCGTCTATGCGCAGCCTATACAC 949
 OY 421 TATGCACACCGTCTGTGTACAGCGTGAAGCTGAAAGCGCACTGCGCGTGG 480
 DB 950 TATGCACACCGTCTGTGTACAGCGTGAAGCTGAAAGCGCACTGCGCGTGG 1009
 OY 481 CTGCGCCCGGAGCGCAAAACCGAGTGTCTTCAATGACAGCGCAAAATCGTTGGT 540
 DB 1010 CTGCGCCCGGAGCGCAAAACCGAGTGTCTTCAATGACAGCGCAAAATCGTTGGT 1069
 OY 541 ATGATGCTGTGCTGCTTTCACACTAGCACTGGAAGAGATCGACAGAAATCGCTGCAA 600
 DB 1070 ATGATGCTGTGCTGCTTTCACACTAGCACTGGAAGAGATCGACAGAAATCGCTGCAA 1129
 OY 601 GAAGCGTATGAGAGATCATCAAGCAATTCCTCCCGCTGAATGCGTGAATCTTGGC 660
 DB 1130 GAAGCGTATGAGAGATCATCAAGCAATTCCTCCCGCTGAATGCGTGAATCTTGGC 1189
 OY 661 ACCAAATTTCTTCAACACCGAGCGCTGTTGCTATGCGTGCGCCCAATGGGTGACTGC 720
 DB 1190 ACCAAATTTCTTCAACACCGAGCGCTGTTGCTATGCGTGCGCCCAATGGGTGACTGC 1249
 OY 721 GGCTGACGTGCTGCTGCTTATGCTTACCTAGGGGCGCATGGCGGCTACGGTGGC 780
 DB 1250 GGCTGACGTGCTGCTTATGCTTATGCTTACCTAGGGGCGCATGGCGGCTACGGTGGC 1309
 OY 781 GGTGCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 DB 1310 GGTGCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1369
 OY 841 TATGTCGCGCAAAACATCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 DB 1370 TATGTCGCGCAAAACATCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429
 OY 901 TATGCAATGCGCGTGGCTGCAACCGAGCTCATGATGTAGAAATTTTGGTACTGAGAAA 960
 DB 1430 TATGCAATGCGCGTGGCTGCAACCGAGCTCATGATGTAGAAATTTTGGTACTGAGAAA 1489
 OY 961 GTGCTTTCTGAAACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

DB 1490 GTGCTTCTGAAACATGACCTGCTGTACGTGATCTTCTGACCTCGCCCATACCGT 1549
 OY 1021 CTGATTCAGATGCTGATCTGCTGACCCGATCTCAAGAAACCGCATACCGTAC 1080
 DB 1550 CTGATTCAGATGCTGATCTGCTGACCCGATCTCAAGAAACCGCATACCGTAC 1609
 OY 1081 TTTGGTGTGAACATTTCCCGTGGGAAAAAACCGCAAGCGCAGCTCTGCGCATGCT 1140
 DB 1610 TTTGGTGTGAACATTTCCCGTGGGAAAAAACCGCAAGCGCAGCTCTGCGCATGCT 1669
 OY 1141 GCCGCTGGAAGTAA 1155
 DB 1670 GCCGCTGGAAGTAA 1684

RESULT 5
 AA209792
 ID AA209792 standard; DNA: 4975 BP.
 XX
 AC AA209792:
 XX
 DT 23-NOV-1999 (first entry)
 DE
 XX
 XX
 KM SAM: S-adenosyl methionine synthase; biotin: bios1; bios2;
 KM bios3; cofactor; decarboxylation; Vitamin H; metk; ss.
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT CDS 530..1684
 FT FT /*tag= a
 FT CDS /*product= "metk"
 FT FT 1782..2987
 FT FT /*tag= b
 FT FT /*product= "bios1"
 PN DE19806872-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 19-FEB-1998: 98DE-1006872.
 XX
 PR 19-FEB-1998: 98DE-1006872.
 XX
 PA (BADI) BASF AG.
 PI Schroeder H:
 PI
 DR WPI: 1999-480095/41.
 DR P-PSDB: AMV33269, AMV33270.
 XX
 PT Production of biotin by expressing S-adenosyl-methionine synthase and
 PT second biotin synthase gene in host cells
 XX
 PS Example 3: Page 36-41; 48pp: German.

CC This invention describes a novel method for the preparation of biotin
 CC (I) which comprises expressing, in a prokaryotic or eukaryotic host
 CC capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
 CC sequence (1), and (b) at least one of the other biotin biosynthesis
 CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
 CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
 CC most animals and some microorganisms. Expression of biotin plus bios1,
 CC bios2 or bios3 significantly increases productivity of biotin
 CC biosynthesis, particularly by at least 3 times. This sequence encodes
 CC the mek/bios1 fusion protein found in plasmid pHS1 which is used in the
 CC method of the invention.
 CC
 CC Sequence 4975 BP: 1198 A: 1350 C: 1329 G: 1098 T: 0 other:


```

Oy 121 GCTTGCAGAACCTACGTAAACCGGCATGGTTTATGTTGGCGCGAATCACCACAGC 180
Db 5144 GCTTGCAGAACCTACGTAAACCGGCATGGTTTATGTTGGCGCGAATCACCACAGC 5203
Oy 181 GCTTGGGTAGACATCGAAGAGATCACCCTTAACACCGCTTCCGAAATTTGGCTATGTGCAT 240
Db 5204 GCTTGGGTAGACATCGAAGAGATCACCCTTAACACCGCTTCCGAAATTTGGCTATGTGCAT 5263
Oy 241 TCCGACATGGGCTTTGACGCTAATCTCTGTGCGGTCTTGACCGCTATCTGGCAACAGCT 300
Db 5264 TCCGACATGGGCTTTGACGCTAATCTCTGTGCGGTCTTGACCGCTATCTGGCAACAGCT 5323
Oy 301 CCTGACATCAACAGAGGGGCTTGACCGCTCCGATCCGCTGGAACAGGGGGGGGTAGACCTAG 360
Db 5324 CCTGACATCAACAGAGGGGCTTGACCGCTCCGATCCGCTGGAACAGGGGGGGGTAGACCTAG 5383
Oy 361 GGTCTGATGTTGGCTACGACACTAATGAACGACGCTGTGATGCCAGCACTATCAC 420
Db 5384 GGTCTGATGTTGGCTACGACACTAATGAACGACGCTGTGATGCCAGCACTATCAC 5443
Oy 421 TATGACACACCGCTGTGACGCTGACGCTGAAGTCCCTAAACCGGACCTCTCCCTGG 480
Db 5444 TATGACACACCGCTGTGACGCTGACGCTGAAGTCCCTAAACCGGACCTCTCCCTGG 5503
Oy 481 CTGGCGCCGAGCGGAAAGCCAGGTGACTTTGATGATGACGACGCAAAATCGTTGGT 540
Db 5504 CTGGCGCCGAGCGGAAAGCCAGGTGACTTTGATGATGACGACGCAAAATCGTTGGT 5563
Oy 541 ATGATGCTGTGCTGCTTCCACTACGACCTGTGAAGATCGACAGCAAAATCGTCCAA 600
Db 5564 ATGATGCTGTGCTGCTTCCACTACGACCTGTGAAGATCGACAGCAAAATCGTCCAA 5623
Oy 601 GAAGCGGTAATGGAAGACATCATCAAGCAATTCGCGCGGTGAATGGCTACCTTCC 660
Db 5624 GAAGCGGTAATGGAAGACATCATCAAGCAATTCGCGCGGTGAATGGCTACCTTCC 5683
Oy 661 ACCAAATCTTCATCAACCCGACCGGCTGTTGCTTATCGGTGGCCCAATGGGTACCTGC 720
Db 5684 ACCAAATCTTCATCAACCCGACCGGCTGTTGCTTATCGGTGGCCCAATGGGTACCTGC 5743
Oy 721 GGTCTGACGTGCTGTAATATATCTGTATACCTAAGCGGCGATGGCCGCTACGCTGGC 780
Db 5744 GGTCTGACGTGCTGTAATATATCTGTATACCTAAGCGGCGATGGCCGCTACGCTGGC 5803
Oy 781 GGTGATCTCTGTGAAGATCGATCAAAAGTGAGCGCTGCCGCTAGCGACAGCT 840
Db 5804 GGTGATCTCTGTGAAGATCGATCAAAAGTGAGCGCTGCCGCTAGCGACAGCT 5863
Oy 841 TATGTCGGAACAAACATGCTGTGCTGCTGGCCGATCGTTGGAATTCAGGTTCC 900
Db 5864 TATGTCGGAACAAACATGCTGTGCTGCTGGCCGATCGTTGGAATTCAGGTTCC 5923
Oy 901 TACGCAATCGCGCTGCTGTAACCCGACCTCATCATGTAGAAACTTTGCTAGTGAAGA 960
Db 5924 TACGCAATCGCGCTGCTGTAACCCGACCTCATCATGTAGAAACTTTGCTAGTGAAGA 5983
Oy 961 GTGCTTGTGAACAACTGCTGCTGTAGTGTGATCTTGCAGCTGCGCCATAGAGCT 1020
Db 5984 GTGCTTGTGAACAACTGCTGCTGTAGTGTGATCTTGCAGCTGCGCCATAGAGCT 6043
Oy 1021 CTGATTCAGATGCTGATCTGCTGACCCGATCTCAAGAAACCGGACGATACGGTAC 1080
Db 6044 CTGATTCAGATGCTGATCTGCTGACCCGATCTCAAGAAACCGGACGATACGGTAC 6103
Oy 1081 TTTGGTCTGAACATTTCCCGTGGGAAAAAACCGCAAGAGCGACGCTCTCGCGCATCT 1140
Db 6104 TTTGGTCTGAACATTTCCCGTGGGAAAAAACCGCAAGAGCGACGCTCTCGCGCATCT 6163
Oy 1141 GCCGCTCTGAAGTAA 1155
Db 6164 GCCGCTCTGAAGTAA 6178

```

```

RESULT 7
ID AAS46261 standard: DNA: 12354 BP.
XX
AC AAS46261:
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding novel mar regulated protein (NIMR) #30.
XX
KW mar regulated polypeptide: NIMR; microbial infection; antibacterial; ds.
XX
OS Escherichia coli.
XX
PN W0200170776-A2.
XX
PD 27-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07478.
XX
PR 10-MAR-2000; 2000US-188362P.
XX
PA (TUFT ) TUFTS COLLEGE.
XX
PI Levy SB, Barbosa TM, Alekshun MN;
XX
DR WPI: 2001-602769/68.
XX
DR P-PSDB: AAU29361.
XX
PT Identifying compounds that modulate a newly identified mar regulated
PT polypeptide activity, useful as antimicrobial compounds. Involves
PT contacting the polypeptide with a test compound.
XX
PS Disclosure: Page 374-382; 526pp; English.
XX
CC The invention relates to a method of identifying compounds that modulate
CC a newly identified mar regulated (NIMR) polypeptide activity. The method
CC comprises contacting an NIMR polypeptide with a test compound under
CC interaction conditions, determining the ability of the compound to
CC modulate the activity or expression of the polypeptide, and selecting the
CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC of microbial infections, and in screening for modulators of NIMR
CC expression and activity. These modulators can be used to reduce the
CC infectivity of a microbe on a surface, and the virulence of a microbe in
CC a subject suffering from an infection. AAS46232; AAS46278 represent
CC Escherichia coli NIMR coding sequences of the invention.
XX
SQ Sequence 12354 BP; 3031 A; 3249 C; 3190 G; 2884 T; 0 other:
Query Match 100.0%; Score 1155; DB 22; Length 12354;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
Oy 1 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGGCGATCTGCAAAATTTGCT 60
Db 5024 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGGCGATCTGCAAAATTTGCT 5083
Oy 61 GACCAATTTTGTGATGCCGTTTGAAGCGGATCTCTGGAACGAGATCCGAAAGCACCGCTT 120
Db 5084 GACCAATTTTGTGATGCCGTTTGAAGCGGATCTCTGGAACGAGATCCGAAAGCACCGCTT 5143
Oy 121 GCTTGCAGAACCTACGTAAACCGGCATGGTTTATGTTGGCGCGAATCACCACAGC 180
Db 5144 GCTTGCAGAACCTACGTAAACCGGCATGGTTTATGTTGGCGCGAATCACCACAGC 5203
Oy 181 GCTTGGGTAGACATCGAAGAGATCACCCTTAACACCGCTTCCGAAATTTGGCTATGTGCAT 240
Db 5204 GCTTGGGTAGACATCGAAGAGATCACCCTTAACACCGCTTCCGAAATTTGGCTATGTGCAT 5263
Oy 241 TCCGACATGGGCTTTGACGCTAATCTCTGTGCGGTCTTGACCGCTATCTGGCAACAGCT 300
Db 5264 TCCGACATGGGCTTTGACGCTAATCTCTGTGCGGTCTTGACCGCTATCTGGCAACAGCT 5323

```

Oy	301	CGTAGCATCAACACGAGGCGGTTGACCGCGGATCCCGCTGGAAACAGGCGCGGGTGACAG	360
Db	5324	CTTACATCAACACAGGCGGCGTTGACCGTCCGGATCCCGTGGACACAGGCGCGGGTGACAG	5383
Oy	361	GGTCGTGATGTTTGGCTACGCACTAATGAACCGACGCTGCTGATGCGCAGCACTATCAC	420
Db	5384	GGTCGTGATGTTTGGCTACGCACTAATGAACCGACGCTGCTGATGCGCAGCACTATCAC	5443
Oy	421	TATTCACACCGCTCTGCTGATACACCGCTCAGGCGTAAGTGGCTATAAAACGGCAGCTTCGCGTGG	480
Db	5444	TATTCACACCGCTCTGCTGATACACCGCTCAGGCGTAAGTGGCTATAAAACGGCAGCTTCGCGTGG	5503
Oy	481	CTGCGCCCGGACCGGAAACACAGAGTCACTTTTCAGTATGACGACGCGCAAAATCGTTGGT	540
Db	5504	CTGCGCCCGGACCGGAAACACAGAGTCACTTTTCAGTATGACGACGCGCAAAATCGTTGGT	5563
Oy	541	ATTCGATCCTGTCTGCTGCTTTCCACTCAGCACTCTGAAAGATCGACACGAAATCGCTGCA	600
Db	5564	ATTCGATCCTGTCTGCTGCTTTCCACTCAGCACTCTGAAAGATCGACACGAAATCGCTGCA	5623
Oy	601	GAAACGGTAATGGAAGAGATCATTAAGCCAAATTCGCCCCGTGAATGGCTGACTTGCC	660
Db	5624	GAAACGGTAATGGAAGAGATCATTAAGCCAAATTCGCCCCGTGAATGGCTGACTTGCC	5683
Oy	661	ACCAAAATTCCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC	720
Db	5684	ACCAAAATTCCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC	5743
Oy	721	GGCTGACTGCTGCTGATAAATTTATCGTGTGATTAACCTAAGCGCGGCAATGGGCGCTCACGGTGGC	780
Db	5744	GGCTGACTGCTGCTGATAAATTTATCGTGTGATTAACCTAAGCGCGGCAATGGGCGCTCACGGTGGC	5803
Oy	781	GGTCGATCTCTCTGCTAAAGATCCATCAAAAGTGAGACCGCTTCCGCGAGCCTACCGACGACGT	840
Db	5804	GGTCGATCTCTCTGCTAAAGATCCATCAAAAGTGAGACCGCTTCCGCGAGCCTACCGACGACGT	5863
Oy	841	TATGTCCCGGAAAAACATTCGTTGCTGCTGGCGCTGGCCGATCGTTGTGAATTTCAAGTTTCC	900
Db	5864	TATGTCCCGGAAAAACATTCGTTGCTGCTGGCGCTGGCCGATCGTTGTGAATTTCAAGTTTCC	5923
Oy	901	TACCAATCGCGCGGCTGGAACGACCTCCATCATGCTGTAGAAACTTTCGGTACTGAGAAA	960
Db	5924	TACCAATCGCGCGGCTGGAACGACCTCCATCATGCTGTAGAAACTTTCGGTACTGAGAAA	5983
Oy	961	GTCGCTTCTGAACAACGACCTGCTGTAAGTGTCTTTCGACCTGCCCCATACGCT	1020
Db	5984	GTCGCTTCTGAACAACGACCTGCTGTAAGTGTCTTTCGACCTGCCCCATACGCT	6043
Oy	1021	CTGATTCAGATGCTGGAATCTGCTCCACCGCATCTACAAAGAAACCGAGGATCCGCTCAC	1080
Db	6044	CTGATTCAGATGCTGGAATCTGCTCCACCGCATCTACAAAGAAACCGAGGATCCGCTCAC	6103
Oy	1081	TTTGGTCTGTGAACATTTTCCCGTGGGAAAAAACGACAAAGGCGAGCTGCTGCGGATGCT	1140
Db	6104	TTTGGTCTGTGAACATTTTCCCGTGGGAAAAAACGACAAAGGCGAGCTGCTGCGGATGCT	6163
Oy	1141	GCCGCTCTGAAGTAA 1155	
Db	6164	GCCGCTCTGAAGTAA 6178	
RESULT 8			
AAS46267			
ID	AAS46267 standard; DNA: 12354 BP.		
XX	AAS46267:		
XX	AC		
XX	18-DEC-2001 (first entry)		
XX	DNA encoding novel mar regulated protein (NIMR) #36.		
XX	mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds		

```

XX Escherichia coli.
PN WO200170776-A2.
PD 27-SEP-2001.
XX
XX 08-MAR-2001; 2001WO-US07478.
XX 10-MAR-2000; 2000US-188362P.
XX
XX (TUFT ) TUFTS COLLEGE.
XX
XX Levy SB, Barbosa TM, Alekshun MN;
PI
DR WPI: 2001-602769/68.
DR P-PSDB: AAU29368.
XX
XX
XX Identifying compounds that modulate a newly identified mar regulated
PT polypeptide activity, useful as antimicrobial compounds, involves
PT contacting the polypeptide with a test compound -
XX
XX
XX Disclosure: Page 424-432; 526pp; English.
XX
XX The invention relates to a method of identifying compounds that modulate
CC a newly identified mar regulated (NIMR) polypeptide activity. The method
CC comprises contacting an NIMR polypeptide with a test compound under
CC interaction conditions, determining the ability of the compound to
CC modulate the activity or expression of the polypeptide, and selecting the
CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC of microbial infections, and in screening for modulators of NIMR
CC expression and activity. These modulators can be used to reduce the
CC infectivity of a microbe on a surface, and the virulence of a microbe in
CC a subject suffering from an infection. AAS46237-AAS46278 represent
CC Escherichia coli NIMR coding sequences of the invention.
XX
XX
SO Sequence 12354 BP; 3031 A; 3249 C; 3190 G; 2884 T; 0 other:

Query Match          100.0%; Score 1155; DB 22; Length 12354;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGCAAAACACCTTTTACGTCGAGTCCGCTCTGAAAGGCATCTGACAAAATTGCT 60
Db 5024 ATGGCAAAACACCTTTTACGTCGAGTCCGCTCTGAAAGGCATCTGACAAAATTGCT 5083
OY 61 GACCAAAATTTTGATGCGCGTTTATAGACCGGATCTCGAAGACAGATCCGAAACACGCGTT 120
Db 5084 GACCAAAATTTTGATGCGCGTTTATAGACCGGATCTCGAAGACAGATCCGAAACACGCGTT 5143
OY 121 GCTTGGCAACCTACTGTAATAAACCGGCATGCTTTAGTTGGCGGGGAAATACACACAGC 180
Db 5144 GCTTGGCAACCTACTGTAATAAACCGGCATGCTTTAGTTGGCGGGGAAATACACACAGC 5203
OY 181 GCCTGGTATGACATGGAAGAGATCACCGGTAACACCGTTCCGGAATTGGTATATGCAAT 240
Db 5204 GCCTGGTATGACATGGAAGAGATCACCGGTAACACCGTTCCGGAATTGGTATATGCAAT 5263
OY 241 TCCGACATGGCGCTTTGACGCTAATCTCTGTGCGGTTCTTACAGCGCTATCGGCAAAACAGTCT 300
Db 5264 TCCGACATGGCGCTTTGACGCTAATCTCTGTGCGGTTCTTACAGCGCTATCGGCAAAACAGTCT 5323
OY 301 CCTGCATCAACACGAGGGGGTTGACCGTCCGATCCGGCTGGAACACAGGGCGGGGTGACAG 360
Db 5324 CCTGCATCAACACGAGGGGGTTGACCGTCCGATCCGGCTGGAACACAGGGCGGGGTGACAG 5383
OY 361 GGTCTGATGTTTGGCTACGCAACTAATGAAACGAGAGTGTATGTCACACACTATACAC 420
Db 5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACGAGAGTGTATGTCACACACTATACAC 5443
OY 421 TATGCACACCGTCTGCTACAGCGCTAAGGCTGAAGTCCGTAATAAGGGCACTTCCCGTGG 480
Db 5444 TATGCACACCGTCTGCTACAGCGCTAAGGCTGAAGTCCGTAATAAGGGCACTTCCCGTGG 5503

```



```

Db 433 TACGGCACCCTCTGCTACACCGCTACGCTAAGCTGCTAAATAATGGACCCCTCCATGG 492
Oy 431 CTGGCCCGGACGCGAAAGCAGTGAATTTTTCAGTATGACGACGCGAAATATGTTGGT 540
Db 433 TTGGCTCCGGATGCAAAAGCAGTCACTTTCCAGTATGACGACGCGAAATATGTTGGT 552
Oy 511 ATTCGATGCTGCTGCTTTTCCAGTACGCTGACGCTGACGACGCTGACGACGCTGACG 600
Db 513 ATTCGATGCTGCTGCTTTTCCAGTACGCTGACGCTGACGACGCTGACGACGCTGACG 612
Oy 601 GAAGCGGATGATGAAGATCATCAAGCAATTCGCGCGGATGATGCTGCTGCTGCTGCTG 660
Db 613 GAAGCGGATGATGAAGATCATCAAGCAATTCGCGCGGATGATGCTGCTGCTGCTGCTG 672
Oy 661 ACCAATTTCTTCATCAACCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 673 ACTAAGTTTTCATCAACCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 732
Oy 721 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 723 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792
Oy 781 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 793 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
Oy 841 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 853 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 912
Oy 901 TACGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 913 TACGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 972
Oy 961 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 973 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032
Oy 1021 CTGATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1033 TTGATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1092
Oy 1081 TTGATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1093 TTGATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152
Oy 1141 GCGGCTGCAAGTAA 1155
Db 1153 GCGGCTGCAAGTAA 1167

```

```

RESULT 10
AAS85725
ID AAS85725 standard; cDNA; 2853 BP.
XX AAS85725:
AC AAS85725:
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #21529.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
RV food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX

```

```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB: ABG21538.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 21529; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS6197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2853 BP; 693 A; 680 C; 791 G; 689 T; 0 other;
XX
XX Query Match 70.6%; Score 815.4; DB 23; Length 2853;
XX Best Local Similarity 89.1%; Pred. No. 1.8e-238;
XX Matches 940; Conservative 0; Mismatches 0; Indels 114; Gaps 1;
XX
Oy 101 AGATCCGAAGACGCGCTTCTGCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 160
Db 1913 AAGATCCGAAGACGCGCTTCTGCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1972
Oy 161 GCGGCGAATACACCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 220
Db 1973 GCGGCGAATACACCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2032
Oy 221 GCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 280
Db 2033 GCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2092
Oy 281 GCGGCTATGCGCAAGAGTCTCTGACATCAACGAGGCGTTGACCGTCCGCTGCTGCTGCTG 340
Db 2093 GCGGCTATGCGCAAGAGTCTCTGACATCAACGAGGCGTTGACCGTCCGCTGCTGCTGCTG 2152
Oy 341 AACAGGCGCGGCGGAGGAGGCTGATGTTTGGCTACGCACTTAATGAACGAGCTGC 400
Db 2153 AACAGGCGCGGCGGAGGAGGCTGATGTTTGGCTACGCACTTAATGAACGAGCTGC 2212
Oy 401 TGATGCGACGACCTATGACCTATGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 460
Db 2213 TGATGCGACGACCTATGACCTATGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2272
Oy 461 AAAACGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 520
Db 2273 AAAACGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2332

```

```

OY 521 ACGACGGCAAAATCGTTGGATGATGATGCTGCTGCTTTTCCACTCAGACTCGAAGAGA 580
    |||||||
Db 233 ACGACGGCAAAATCGTTGGATGATGATGCTGCTGCTTTTCCACTCAGACTCGAAGAGA 2392
OY 581 TCGACCGAAGAAATCGCTGCAAGAACCGGTAATGAGAGATCATCAACCAATTCCTCCCG 640
    |||||||
Db 2393 TCGACCGAAGAAATCGCTGCAAGAACCGGTAATGAGAGATCATCAACCAATTCCTCCCG 2452
OY 641 CTGAATGCGTACCTTCCGACCAAAATCTTCATCAACCGGCGGCTTGGTTATCG 700
    |||||||
Db 2453 CTGAATGCGTACCTTCCGACCAAAATCTTCATCAACCGGCGGCTTGGTTATCG 2496
OY 701 GTGGCCCAATGGGTGACTGGGCTCTGACTGCTGTAATAATTATGATACCTACGGCG 760
    |||||||
Db 2497 ----- 2496
OY 761 GCATGCGCGCTACAGCGTGGCGGTCATCTCTGTAAGATCCATCAAAAGTGACCGTT 820
    |||||||
Db 2497 -----GATCATCAAAAGTGACCGTT 2518
OY 821 CCGCAGCCTACGAGCAGCTTATGCGGAAAAAATCCTTGGTGGCGGCTGGCCGATC 880
    |||||||
Db 2519 CCGCAGCCTACGAGCAGCTTATGCGGAAAAAATCCTTGGTGGCGGCTGGCCGATC 2578
OY 881 GTTGTGAATTCAGGTTTCCCTACGCAATCGCGTGGCTGGAACCGACCTCCATCATGCTAG 940
    |||||||
Db 2579 GTTGTGAATTCAGGTTTCCCTACGCAATCGCGTGGCTGGAACCGACCTCCATCATGCTAG 2638
OY 941 AAATTTCCGTACTGAGAAAGTGCCTTCTGACACACTGACCCCTGCTGGTACGTGACTTCT 1000
    |||||||
Db 2639 AAATTTCCGTACTGAGAAAGTGCCTTCTGACACACTGACCCCTGCTGGTACGTGACTTCT 2698
OY 1001 TCGACCTGGCGCCATACGGTCTGATTCAGATGCTGGATGCTGGTGCACCGCATCTACAAG 1060
    |||||||
Db 2699 TCGACCTGGCGCCATACGGTCTGATTCAGATGCTGGATGCTGGTGCACCGCATCTACAAG 2758
OY 1061 AAACCCGACGATACGCTACTTTGTCGTGAGACATTTCCCGTGGGAAAAAACGACAAAG 1120
    |||||||
Db 2759 AAACCCGACGATACGCTACTTTGTCGTGAGACATTTCCCGTGGGAAAAAACGACAAAG 2818
OY 1121 CGCAGCTGCTGGCGGATGCTGGCGGCTGTGAAGTAA 1155
    |||||||
Db 2819 CGCAGCTGCTGGCGGATGCTGGCGGCTGTGAAGTAA 2853

RESULT 11
AAS3462
ID AAS3462 standard; DNA: 1155 BP.
XX
AC AAS3462:
XX
DT 13-FEB-2002 (first entry)
XX
DE Haemophilus influenzae DNA for cellular proliferation protein #244.
XX
KM Antisense: ds: prokaryotic cellular proliferation gene:
XX
KW antibiotic: antibacterial: drug design.
XX
OS Haemophilus influenzae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001: 2001MO-US09180.
XX
PR 21-MAR-2000: 2000US-191078P.
PR 23-MAY-2000: 2000US-206848P.
PR 26-MAY-2000: 2000US-207272P.
PR 23-OCT-2000: 2000US-242578P.
PR 27-NOV-2000: 2000US-253625P.
PR 22-DEC-2000: 2000US-257931P.
PR 16-FEB-2001: 2001US-269308P.

```

```

XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ:
PI Yamamoto RT, Xu HH;
XX
DR MPI: 2001-61495/70.
XX
DR P-PSDB: A4U35603.
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
XX Claim 27: Seq ID No 7099; 511pp: English.
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/publshd_pcl_sequences.
XX
SQ Sequence 1155 BP; 348 A; 190 C; 267 G; 350 T; 0 other:
Query Match 55.5%; Score 641.4; DB 23: Length 1155;
Best Local Similarity 72.2%; Pred. No. 1,8e-185;
Matches 834; Conservative 0; Mismatches 321; Indels 0; Gaps 0:
OY 1 ATGCAAAACACCTTTTACGTCGAGTCGCTTGAAGAGCATCTGACAAATTCGT 60
    |||
Db 1 ATGCTAGTATTTATTTATTTACATCTGAATCAGTCTGAAAGCGATCAATAAATTGGC 60
OY 61 GACCAATTTCTGATGCGGCTTTAGACGCGATCTCGAAGCATCGGAAGCAGCGTT 120
    |||
Db 61 GATCAAAATTTCTGATGCGGCTTTAGTGAATCTTAACCAAGATCAAAAGCAGCAGTA 120
OY 121 GCTTGGCAAACTACGTAATAAACCGGCAATGTTTGAATTGGCGCAATACACAGC 180
    |||
Db 121 GCTTGGCAAACTACGTAATAAACCGGCAATGTTTGAATTGGCGCAATACACAGC 180
OY 181 GCGTGGGTAGACATCGAAGATCACCGCTTACACCGCTTGCAGAAATTCGCTATGTCAT 240
    |||
Db 181 GCAATGGGTATATGGAATTTAACCCCTAAAGTGAATTTGATATGATGTTATGAACAT 240
OY 241 TCCGACATGGCGTTTGAACGCTAATCCTGCGGCTTCTGAGCCCTATCGGCAACAGTCT 300
    |||
Db 241 TCTGAATAGGCGCTTTGATGCTCATTTCTTGTCAGTCTTATGCGATGGTAAACATCT 300
OY 301 CCGACATCAACGAGCGCTTGAACCGTGCATCCGCTGGAACGAGCGGCTGACAG 360
    |||
Db 301 GCAATATTTAATCAAGCGCTTGAATCGTGAATAATTCATTAACAGCGGAGCGACCA 360
OY 361 GGTCTGATGTTTGGCTACGCACTAATGAACGAGCGTGTATGCCAGACTATTCACC 420
    |||
Db 361 GGTATATGTTTGGTATGTCACCAATGAACGAGATGATTAATGCTGAGCATTAATC 420
OY 421 TATGCAACAGCTGTGTAAGAGCGTCAAGAGTGGTAAAGAGCGGCACTGCGCTG 480
    |||
Db 421 TATGCAACATGCTTAAATGGAAGAAACATCTGAAGTGGCTTAAAGCGGTAATTAAGCTG 480
OY 481 CTGCGCGCGGACCGCAAAACCGAGGTGACTTTTCAGTATGACGACGCAAAATCTGTTG 540

```

```

Db 441 TTACCCAGATGCGAAAGCCAGTTACTTAAATATGAGATATATATATGTTGGT 540
Oy 541 ATGATGCTGTGCTTCTTCCATCAGCAGCTGAGAGATCGACAGAAATCCCTGCA 600
Db 541 GTGATGCGGTTGCTTCTTACTCAACATCTCTGAGAGATTAGCAAAAGATTACAT 600
Oy 601 GACGCGTAATGAAGAGATCATCAGCCAAATTCGCCCGCTGAATGGCTACTTCTGCC 660
Db 601 GAAGCTGTATGAGAGAAATATATCAACCTGTATCCAACTGATGATGCTTCTTAAGAA 660
Oy 661 ACCAAATCTTCATCAACCCAGCGGTGTTGCTTATCGGCGCCCAATGGGACAGTGC 720
Db 661 ACAAAATCTTCATTAACCCAGCTGTGTTGTTTATCGGCGCCCAATGGGACAGTGT 720
Oy 721 GCTGCTGCTGCTGAATATATCTGTATGATCCTACGCGGCGATGGCGCTACGCTGCC 780
Db 721 GGTTAATGCGCCGTAATATCATGCTGATCTTACGCTGGTCCGCGCTGACGCTGGT 780
Oy 781 GGTGCTATCTCTGTAAGATCATCAAAAGTGACCGCTTCCGACGCTTACGACGCT 840
Db 781 GCGCATCTCAGGAAAGAACATCTAAAGTTGACCGCTTCTGCGGCTTATGCTGCACGT 840
Oy 841 TATGCGGGAACATGCTGCTGCTGCGCTGCGCGATCGTTGGAATTCAGGTTCC 900
Db 841 TATGCGGGAACATGCTGCTGCTGCGCGATCGTTGGAATTCAGGTTCT 900
Oy 901 TACGCAATGCGCTGCTGCTGCAACCGACCTCATCATGATAGAAATTCGCTACTGAGAA 960
Db 901 TATGCAATGCTGCTGCTGCAACCGACCTCATCATGATAGAAATTCGCTACTGAGAA 960
Oy 961 GTGCTTTGCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 961 GTGCTTTGCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Oy 1021 CTGATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 TTAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Oy 1081 TTTGCTGCTGCAACATTTTCCGTTGGGAAAAACGCAAAAGCGCAGCTGCTGCTGCT 1140
Db 1081 TTTGCTGCTGCAACATTTTCCGTTGGGAAAAAGTGATGCTGCTGCTGCTGCTGCT 1140
Oy 1141 GCGGCTGCAAGTAA 1155
Db 1141 GCAGCTTAAATAA 1155

```

RESULT 12

AAS54057 standard; DNA: 1191 BP.

```

AC AAS54057;
DE 13-FEB-2002 (first entry)
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #188.
KW Antisense; ds: prokaryotic cellular proliferation gene;
KW Antibiotic; antibacterial; drug design.
OS Pseudomonas aeruginosa.
XX WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207272P.
XX PR 23-OCT-2000; 2000US-242578P.

```

```

PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
PA (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlson KL, Zykkind JW, Wall D, Traxwick JD, Carr CJ;
PI Yamamoto RT, Xu HH;
DR WPI: 2001-611495/70.
DR P-PSDB: AAU36198.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27: Seq ID No 7694; 511bp; English.
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1191 BP: 222 A; 420 C; 365 G; 184 T; 0 other:

```

Query Match 50.8%; Score 586.2; DB 23; Length 1191;
 Best Local Similarity 71.2%; Pred. No. 1,4e-168;
 Matches 774; Conservative 0; Mismatches 313; Indels 0; Caps 0;

```

Oy 12 CCTTTTACGTCGAGTCCTCTCTGAAAGGCACTCTGCAAAATTCGTCGCAATTTTC 71
Db 15 GCTTTTACGTCGAGTCCTCTCTGAAAGGCACTCTGCAAAATTCGTCGCAATTTTC 74
Oy 72 TGATGCGGCTTTTACGTCGAGTCCTCTCTGAAAGGCACTCTGCAAAATTCGTCGCA 131
Db 75 CGACCGGCTGTCGAGTCCTCTCTGAAAGGCACTCTGCAAAATTCGTCGCAATTTTC 134
Oy 132 CTACGTAATAAAGCGCATGTTTATGTTT3CGGCGGAATTCACACAGCGCTGCTGCA 191
Db 135 CCGTGTCAAGACCGCGCGGATGATCATGCGCGGATGATCATGCGCGGATGATCATG 194
Oy 192 CATGCAAGATGATCAACCGCTTACACCGCTT3CGGAATTCGCTATGCTTCCGACATGG 251
Db 195 CCGTGTCAAGACCGCGCGGATGATCATGCGCGGATGATCATGCGCGGATGATCATG 254
Oy 252 CTTGAGGCTTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 311
Db 255 CTTGAGGCTTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 314
Oy 312 CCAGGCGCTTGACCGCTGCGCTGCGTGAACAGAGCGCGGCTGACCAAGGCTGATGTT 371
Db 315 CCAGGCGCTTGACCGCGCGGATGATCATGCGCGGATGATCATGCGCGGATGATCATG 374
Oy 372 TGGCTACCACTATGAACCGGACGTCGTCATGCGACGACCTATGACCTATGACACCG 431
Db 375 GCGCTACCGCGACGACGACGACGTCGTCATGCGCGGCTGCTGCTTCTCTCCACCG 434
Oy 432 TCTGCTACAGCGCTGAGTGAAGTGCATGAAGGCGCACTGCGCGTGCAGCGCCGCA 491

```



```
Db 435 CTTGTCGAGCCGACGCGGAGCGGCAAGTCCGGCTGTCCTGCTGCTGCGCCGGA 494
Oy 492 CCGCAAGAGCAGTATTTTACATGATGACGCAAAATGTTGATGATGCTGT 551
Db 495 CCGCAAGTCCAGTACCTGCCGCTACGAGGCGGCAAGTGGTGGCATGATCGGT 554
Oy 552 CGTGTCTTCCACTACGACTCTGAGAGATGACGCAAAATGCTGCAAGAACGGTAAT 611
Db 555 GGTGCTGTCACCCAGCAACCCGGAAGTCTCTATAGGACCTGCGGACGGCGGAT 614
Oy 612 GGAAGATCATCAACCCATTTGCGCGCTGATGCTGACTTCTGCCACCAATTTCT 671
Db 615 GGGAGTCACTATCAACAGAGTGTGCGGCGGAGCTGTCACAGAGACCCAGTTCCA 674
Oy 672 CATCAACCCGACGGTCTGTTGCTATGCGTGGCCCAATGGTGAGTGGCGTCTGACTGG 721
Db 675 CATCAACCCGACGGCACTTCTGATGCGGCGCGGCTGCGGACTGCGGCTGACCGG 734
Oy 732 TCGTAAATTTATCTGATACCTACGCGGCGCATGCGGCGTACGGTGGCGTCAATCTC 791
Db 735 ACCCAAGATCATGCTGAGCTCTACGCGGCGCATGCGGCGGACGGCGGCGCGTTCTC 794
Oy 792 TCGTAAAGATCCATCAAAAGTGGACCGTTCGCGAGCTACGACGACGTTATCTCCGAA 851
Db 795 CCGCAAGGACCCGTCACAGGTGACCGCTCCGCGCTACGCGCGCGCTACGTTGCGGAA 854
Oy 852 AAACATCGTTGCTGCTGCGGCGCGCATGCTGTAATTCAGTTTCTTCCAGCAATCGG 911
Db 855 GAACATCTGTCGCGCGCGCGCTGCGGCGGCTGCGGAGATTCAGGTTTCTTCCAGCCATCGG 914
Oy 912 CGTGTGTAACGACCTTCATCATGTTAGAAATTTGCGTACTGAGAAATGCGCTTCTGA 971
Db 915 CGTGCCCAAGCGGACCTTCATCATCAACCTTCGCGACGCGGCAAGTGCAGCGACA 974
Oy 972 ACAACTGACCTCTGCTGCTGATGATGTTCTTGCACCTGCGCCCATACGGTCTGATTAGAT 1031
Db 975 GAAGATCGTGCAACTGTGTCGCGGAGCACTTGCACCTGCGCCCTACGCGATCCAGAT 1034
Oy 1032 GGTGATCTGTCGACCCGATCTACAAAGAACCGACGATAGGTCACCTTGTGCTGTA 1091
Db 1035 GCTCGACCTGCTTCACCCGATGTACCAAGCGGACCGCGCTACGGCCACTTGGTGTCTCA 1094
Oy 1092 ACATTTTC 1098
Db 1095 TCCGTTTC 1101

RESULT 13
AAS90256
ID AAS90256 standard: cDNA: 2577 BP.
XX
AC AAS90256:
XX
XX 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagenetic protein #26060.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
```

```
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB: ABC26069.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX
PS Claim 1: SEQ ID No 26060: 103pp: English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
Oy Sequence 2577 BP: 629 A: 676 C: 708 G: 564 T: 0 other:
XX
XX Query Match 46.2%; Score 533.4; DB 23; Length 2577;
XX Best Local Similarity 99.8%; Pred. No. 2,8e-152;
XX Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 543 CGATGCTGTCGCTTCCATCAGCAGCTCTGAGAGATGACACCAAAATCGTGCAGAA 602
Db 549 CGATGCTGTCGCTTCCATCAGCAGCTCTGAGAGATGACACCAAAATCGTGCAGAA 608
Oy 603 ACCGTAATGAGAGATCATCAAGCAATTCGCGCGGTAATGCTGACTTCTCCAC 662
Db 609 ACCGTAATGAGAGATCATCAAGCAATTCGCGCGGTAATGCTGACTTCTCCAC 668
Oy 663 CAAATTTCTCATCAACCCGACCGGCTTTCGTTATCGGTGCGCCAAATGGGTACTGCGG 722
Db 669 CAAATTTCTCATCAACCCGACCGGCTTTCGTTATCGGTGCGCCAAATGGGTACTGCGG 728
Oy 723 TCTGACTGTCGTAATTAATTCGTTATCTGATACCTAGCGCGGATGCGCGTACGGTGGCG 782
Db 729 TCTGACTGTCGTAATTAATTCGTTATCTGATACCTAGCGCGGATGCGCGTACGGTGGCG 788
Oy 783 TGCATCTCTGTAAGATCATCAAAAGTGAAGCGGCTCGGAGCTTACGACACGCTTA 842
Db 789 TGCATCTCTGTAAGATCATCAAAAGTGAAGCGGCTCGGAGCTTACGACACGCTTA 848
Oy 843 TGTGCGAANAACATGCTGCTGCTGCGCTGCGGATCGTTGTGAATTCAGGTTTCCTA 902
Db 849 TGTGCGAANAACATGCTGCTGCTGCGGATCGTTGTGAATTCAGGTTTCCTA 908
Oy 903 CCGAATCGCGGCTGCTGAGACCGACCTTCATCATGTGAGAAATCTTGGGTACTAGAAAT 962
Db 909 CCGAATCGCGGCTGCTGAGACCGACCTTCATCATGTGAGAAATCTTGGGTACTAGAAAT 968
Oy 963 GCCTTCGAACAACACTACCTGCTGTAGTGTCTTCTGACCTGCGCCCATACGCTCT 1022
Db 969 GCCTTCGAACAACACTACCTGCTGTAGTGTCTTCTGACCTGCGCCCATACGCTCT 1028
Oy 1023 GATTCAAGATGCTGATCTGCTGCACCCGATCTTACAAAGAACCGAGCATACGCT 1077
```


PI Relinkainen T., Nyssöelae A., Kerovuori J.
 XX MPI: 2000-224686/19.
 DR P-PSDB: AAY70040, AAY70041, AAY70042.
 XX
 XX New methyltransferases, useful for e.g. producing transgenic plants
 PT with increased stress tolerance, pathogen resistance or nutritional
 PT value as animal feed or for improving the viability of microorganisms
 PT in the food industry
 PS
 PS Claim 1: Figure 8: 176pp; English.

XX The patent discloses the use of novel methyltransferases catalysing one
 CC of the following reactions: glycine to sarcosine (GSM), sarcosine to
 CC dimethyl glycine and dimethyl glycine to betaine (SDMT) or an enzyme
 CC catalysing the synthesis of S-adenosyl methionine (SAM). The present
 CC sequence is the Ectothiorhodospira halochloris betaine operon encoding
 CC the methyltransferases and SAMs, which can be used to increase
 CC intracellular concentrations of these compounds. The polynucleotides can
 CC be used to increase salt, freezing or cold tolerance, increase
 CC resistance to drought stress, pathogens or induce pathogenesis-related
 CC proteins in plants. Transgenic organisms can be used as an animal feed
 CC ingredient. The polynucleotides can also be used to enhance pH tolerance
 CC and improve viability of organisms when subjected to environmental
 CC stress. They can decrease inclusion body formation when used in
 CC conjunction with polynucleotides encoding a heterologous protein.
 CC
 XX Sequence 3219 BP: 745 A: 844 C: 937 G: 693 T: 0 other:

Query Match 43.8%; Score 505.8; DB 21: Length 3219;
 Best Local Similarity 71.3%; Pred. No. 8.7e-144;
 Matches 682: Conservative 0; Mismatches 272; Indels 3; Gaps 1:

Oy 9 ACACCTTTTACGCTGCGTCTGAGGCGATCTGACAAATTTGCGACCAAT 68
 Db 2017 ATATCTATTACCTGTGATGCTGTGAGGCCACCGGACAAATGGCCGACCAAT 2076
 Oy 69 TTCTGATGCGCTTTAGACGCGATCTCGAACAGATCCGAAAGCAGCGCTTGTGCA 128
 Db 2077 TTGCGATGACATGCTGATGATGATGCTGCGGAGATCCAAAGTCCGGGTGTGCA 2136
 Oy 129 AACCTAAGTAAACCGGATGTTTGTGCGGCGCAATCACCACGCGCTGCT 188
 Db 2137 GACGATGATTCAGACCGCGATGTTGTCGCGCAATTAAGACCAATGCCAAGAT 2196
 Oy 189 AGACATCGAAGATCACCGCTGACACCGTTCGCAATTTGCTATGTCATCCGACAT 248
 Db 2197 CATCTGTGCGCGTGTGATGAGAGTGTCCGATATCGCTACACCAAGCTCAGATAT 2256
 Oy 249 GCGCTTTGACGCTACTCTGTGCGTTCTGAGCCCTATCGCAAAACAGTCTCTGACAT 308
 Db 2257 GCGCTTTGATCCGACACTGTGCGGTACTCAACGCCCTCGCGGAGCAATCCCGGACAT 2316
 Oy 309 CACACGCGCTTGAACGCTGCGGATCCGCTGACACAGGCGCGGTGACACAGGCTGTGAT 368
 Db 2317 CAATCAAGCGCTTGAACGCGGAGAGAGAGAGAGAGAGGCGCGGTGACCAAGACTGAT 2376
 Oy 369 GTTGGCTACGCACTATGAAACGAGCTGTGATGCCAGACCTATCACCCTATGACACA 428
 Db 2377 GTTTCGTTAGCGCACCAATGAGACGAGCTCTCATGCGCGCAGCATCACTACTCGCA 2436
 Oy 429 CCGTGTGACAGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 488
 Db 2437 CCGTGTGTTAGCGCGCAATCGAAGTCCGTACTCAGAAAGCTGCGGTGTGCGGCC 2496
 Oy 489 GGACGCGAAAGCGAGTACTTTTCAATGACGAGCGCAAAATCGTTGTATGATGAC 548
 Db 2497 GGACGCTAAGACCGAGTCACTTCAAGTACGAGCGATAGATGTCGTTGCGATGC 2556
 Oy 549 TGTCTGCTTTGCACTCAGCACTGGAAGATTCACCAAGAAATCGCTGCAAGAAGCGT 608
 Db 2557 GGTGTGCTTCTCCACGACGATGACGAGACCGTTGATCAGAGACGTTCCACGAGGCGT 2616

Oy 609 AATGAGAGATCATCAAGCCAAATTTCT--GCCCGCTGAATGGCTGACTTCTGCCACCA 665
 Db 2617 GATGAGAGATCATCAAGCCCAATCTCTGATGACACCGGCTGGGTGACCAAGACCA 2676
 Oy 666 ATTCTTCAACCCGACCGGCTGTTGTATGCTGCGCCAAATGGGTACTGGGCT 725
 Db 2677 GTACCACTTAACCCGACCGGCTGTTGTATGCTGCGCCAAATGGGTACTGGGCT 736
 Oy 726 GACTGCTGTAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
 Db 2737 GACCGCGCTAAGATCATCTGCTGATGATGATGATGATGATGATGATGATGATGAT 2796
 Oy 786 ATTCTTCTGAAGATCATCAAGCCAAATTTCT--GCCCGCTGAATGGCTGACTTCTGCCACCA 845
 Db 2797 GTTCTCCGCAAGATCATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2856
 Oy 846 CGGCAAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 905
 Db 2857 AGCTAAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2916
 Oy 906 AATCGCGTGGCTGAACCGACCTCCATCATGATGATGATGATGATGATGATGATGAT 962
 Db 2917 CATCGGTGTGCGGACCGACCTCGGTATGATGATGATGATGATGATGATGATGATGAT 2973

Search completed: April 22, 2003, 10:11:47
 Job time : 351.095 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:19:24 ; Search time 50.3875 Seconds

(without alignments)
7029.746 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaacacccctttac.....atgcgcgcgcgcgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/6A.COMB.seq:*
4: /cgn2_6/pdata/1/lna/6B.COMB.seq:*
5: /cgn2_6/pdata/1/lna/CTUS.COMB.seq:*
6: /cgn2_6/pdata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	374	32.4	4848	US-08-955-957A-1	Sequence 1, Appl
2	374	32.4	4848	US-08-955-957A-4	Sequence 4, Appl
3	374	32.4	4848	US-08-955-957A-6	Sequence 6, Appl
4	369.8	32.0	1208	US-08-403-852D-4	Sequence 4, Appl
5	369.8	32.0	1208	US-08-510-646B-4	Sequence 4, Appl
6	369.8	32.0	1208	US-09-231-818-4	Sequence 4, Appl
7	369.8	32.0	5392	US-08-403-852D-1	Sequence 1, Appl
8	369.8	32.0	5392	US-08-510-646B-1	Sequence 1, Appl
9	369.8	32.0	5392	US-09-231-818-1	Sequence 1, Appl
10	368.2	31.9	1191	US-09-273-686-1	Sequence 1, Appl
11	368.2	31.9	8494	US-08-961-527-163	Sequence 163, App
12	349.2	30.2	1248	US-09-134-001C-2240	Sequence 2240, Ap
13	332.2	28.8	1283	US-09-318-448-22	Sequence 22, Appl
14	316	27.4	3228	US-09-318-448-20	Sequence 20, Appl
15	313.2	27.1	4403765	US-09-103-840A-2	Sequence 2, Appl
16	313.2	27.1	4411529	US-09-103-840A-1	Sequence 1, Appl
17	297.6	25.8	1693	US-09-320-878-23	Sequence 23, Appl
18	208	18.0	2774	US-09-363-189B-3	Sequence 5, Appl
19	169	14.6	10061	US-09-221-017B-672	Sequence 672, App
20	159.4	13.8	725	US-08-998-416-1070	Sequence 1070, Ap
21	35.4	3.1	50937	US-09-428-517-1	Sequence 1, Appl
22	34.8	3.0	25165	US-09-453-702B-39	Sequence 39, Appl
23	32.8	2.8	423	US-08-470-179-182	Sequence 182, App
24	31.8	2.8	5395	US-09-221-017B-27	Sequence 27, Appl
25	31.8	2.8	10061	US-09-221-017B-672	Sequence 672, Appl
26	31.4	2.7	534	US-09-081-180-20	Sequence 20, Appl
27	31.4	2.7	534	US-09-040-786-20	Sequence 20, Appl

C 28	31.4	2.7	1439	4	US-09-056-556-167	Sequence 167, App
C 29	31.4	2.7	1439	4	US-09-072-596-162	Sequence 162, App
C 30	31.4	2.7	25165	4	US-09-453-702B-39	Sequence 39, Appl
C 31	31.4	2.7	30001	2	US-08-125-468-1	Sequence 1, Appl
C 32	31.4	2.7	30001	2	US-08-474-933-1	Sequence 1, Appl
C 33	31	2.7	3616	4	US-09-212-247C-7	Sequence 7, Appl
C 34	30.8	2.7	10216	2	US-08-875-154-1	Sequence 1, Appl
C 35	30.8	2.7	19182	2	US-08-850-880-11	Sequence 11, Appl
C 36	30.8	2.7	19182	2	US-08-944-916-11	Sequence 11, Appl
C 37	30.8	2.7	19182	4	US-09-272-432A-11	Sequence 11, Appl
C 38	30.2	2.6	1434	4	US-09-491-772-2	Sequence 2, Appl
C 39	30.2	2.6	1632	1	US-08-362-232-1	Sequence 1, Appl
C 40	30.2	2.6	1632	1	US-08-814-196-1	Sequence 1, Appl
C 41	30.2	2.6	6914	1	US-08-920-812-22	Sequence 22, Appl
C 42	30.2	2.6	6914	1	US-08-920-837-22	Sequence 22, Appl
C 43	30.2	2.6	6914	1	US-08-921-177-22	Sequence 22, Appl
C 44	30.2	2.6	6914	1	US-08-362-577C-22	Sequence 22, Appl
C 45	30.2	2.6	6914	2	US-08-920-828-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-955-957A-1
Sequence 1, Application US/08955957A
Patent No. 6312920
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39, 872
REFERENCE/DOCKET NUMBER: P-10162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4848 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 986..2209
FEATURE:
NAME/KEY: CDS
LOCATION: 2241..3341
US-08-955-957A-1
Query Match 32.4%; Score 374; DB 4;
Best Local Similarity 59.7%; Pred. No. 4.2e+101;
Length 4848;

Matches	728:	Conservative	0:	Mismatches	420:	Indels	72:	Gaps	3
Oy	1	ATGCGAAAACACCTTTTACGTCGCGAGTCGCGCTCTGTGAAGGGCATCTTGACAAATGCT							60
Dh	916	ATGTACAGCTCGCGTGTACACCTCGGAGTCCCGTAGACCGAGGGCCACCCCAAGAATCGCC							1045
Oy	61	GACCAAAATTTTCTGATCCCGCTTTTAGACCGGATCCCTCGAAGAGATCCGAACACGCGCTT							120
Dh	1046	GACCGGATCAGCGAACCGTCTCGATCCCTCGCTCGCCCGGAGACCCTGGGCGACGAGTTC							1105
Oy	121	GCTTTCGAAACCTACGTAAAAACCGGCAATGGTTTTAGTTGGGGGGAATTCACACACG							180
Dh	1106	GCGGTCGAGAGCCCTGATATCACACCGCGTCAAGTCCATATCGCCGGGAGGTCAACACACG							1165
Oy	181	GCTGGGTGAGCATCGAAGAGATCACCCGTACACCGCTTCGCGAAATTTGGCTATGTGAT							240
Dh	1166	GCGTACGGGCCCATCGGCCCACTGGTGGCGGACACCGTCTGTCCATCGATACGACTCC							1225
Oy	241	TCCGACATGGGCTTTGACGGCTACTCTCTGTGGGTTCTTGAGGGCTATCGGCAACAGTCT							300
Dh	1226	TCGGCAAGGGCTTTCGACGGCGGCTCTGTGGGGGTGTGCTGTCCATCGGCGCGAGTCC							1285
Oy	301	CGTCGATCAACACGAGCGGTTGAC-----							324
Dh	1286	CCGGACATCGCCCGGGGTGTGGACACCCCGTACGAGCGGGGGCGGGGGGACAGGCCCGG							1345
Oy	325	-----CGTGCCGATCCGCTGGACAGAGGGCGGGGTGACAGGAGGTCATGTTTGGC							375
Dh	1346	GCGCGACGGGTGACGACTGGACCGGAGGCGGGGAGGACAGGCGCTGATGTTTCGGC							1405
Oy	376	TACGCACTAATGAACCGGACTGCTGATGGCAGACCATATACCTATATGACACGCTCG							435
Dh	1406	TACGCTTCGACGACGACACCCCGAGCTGATGCGGTCGCCATACCTTGGCGGACCGGCTC							1465
Oy	436	GTACAGCTCAAGGCTGGAAGTGGCTAAAAACGGCAGCTCTCGGTGGCTGGCGGAGCGG							495
Dh	1466	TCCCGGCGGCTGTCCGAGAGTGGCGAAGAGGACAGATCCCTACTCCGCCCGACGGC							1525
Oy	496	AAAACCGAGTGACTTTTTCAGTATGACGACGCGCAAAATGTTGGATATCGATGCTGTGTC							555
Dh	1526	AAGACCCAGGTCAACCATGATGATGACGAGGCGGACAGGCGGTCCGCTCGACAGGGTCGG							1585
Oy	556	CTTTCTACTCAGACACTCTGAGAGATCCGACCGAAGAAATCGCTCTGACAGAGCGGTAATGG							615
Dh	1586	GTCTCTCTCCACAGCAGCGCTCCGGAATCGACCTCGGACTCCCTACTGGCGGCCACATCCCG							1645
Oy	616	GAGATCATCAAGCCAAATCTGCCGCGTGAATGGCTGACATTGCGCCACAAATCTTCA--							673
Dh	1646	CGGCAATGTGTGGAAGCCCGCTCTCCGCGGACGTGGCGGAGAGCGGATCAAGCTGGACACC							1705
Oy	674	-----TCAACCGGACCGGCTGTTTGTATCGTGGGCCCAATGGGT							714
Dh	1706	GCGGCTACCGGCTGCTGTGTTCAATCCACCGGCGGGTTTGAATGGGGCGGATGGGG							1765
Oy	715	GACCTCGCTCTGACTGTGTCTAAATTTATCGTTGATACCTACGCGCGCATGGCGCTCAC							774
Dh	1766	GACGGGGGCTGACCGGGCGGCAAGATCATCATCTGACACTACGGGCGCATGTGCGCGCAT							1825
Oy	775	GTTGGCGGTGCACTTCTGTGTAAAGTCTCATAAAGGTGAGACGTTCCGACAGCTTACGGA							834
Dh	1826	GCGGCGGGTCTTCTTCGGGCAAGGACCCGTCCAAGCTGGACCGTTTGGCGCGGTACGGA							1885
Oy	835	GCACGTTATGTGCGCAAAACATCGTTGTCTGTGGCTGGCGGATCGTTGTGAATAATCAG							894
Dh	1886	ATGCGCTGGGTGGCAAGAACGTCGTGGGGGGGGGCGGCTCGCGGTGTGAGGTGCAC							1945
Oy	895	GTTTCTACGCAATCGCGGTGGCTGAACCGACGACTCCATCATGGTAGAAACTTTGCGTACT							954
Dh	1946	GTCGCTACAGCCATCGGCAAGGCGGACCGGTCGCTGTTGTGTGAAGACCTTCGCGCAC							2005
Oy	955	GAGAAAGTCCCTTTGAAACAACACTGACCCCTGCTGTGATCGTAGTCTTCGACCTCGGCCCA							1014
Dh	2006	CGCACCTGTGACGTGAAGGATCGAGCGAGGAGATCGGCGAGAGTCTTTCACCTTCGCCCG							2065

```

Oy 1015 TAGGCTGATTCAGATGCTGATGATGCTGTCACCCGATCTACAAAGAACCCGACATAC 1074
      || || || || || || || || || || || || || || || || || || || || ||
Db 2066 GCCCGGATCATCTCGGGATCTGGACCTGCTCCGCCGATCTACGCCAACCCGCCCTAC 2125
Oy 1075 GATCAGCTTGAGTCGTGAC-----ATTCCCGTGGAAAAACCGACAAAGCCGACCTG 1128
      || || || || || || || || || || || || || || || || || || || || ||
Db 2126 GCCCAGCTTCGGCCGGGAATGCGCCAGATTCACTGGGAACGACACCGACCGACGACG 2185
Oy 1129 CTGCGGATGCTGCCGCT 1148
      || || || || || || ||
Db 2186 CTCATCGCCGCGCGCGCT 2205

RESULT 2
US-08-955-957A-4
; Sequence 4, Application US/08955957A
; Patent No. 6312920
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Rostock Jr., Paul R.
; TITLE OF INVENTION: SAM Operon
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0., Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955.957A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: P-10162
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4848 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3338..4255
; US-08-955-957A-4

Query Match 32.4%; Score 374; DB 4; Length 4848;
Best Local Similarity 59.7%; Pred. No. 4.2e-101;
Matches 728; Conservative 0; Mismatches 420; Indels 72; Gaps 3;
Oy 1 ATGCGAAAAACACCTTTTATCGTCGAGTCCTGTCGAAAGGCGCATCTGCAAAAATTGCT 60
Db 986 ATGTATCCGCTGCTTTCACCTCGGAGTCTGTGACCCGAGGCGCACCAAGATCCG 1045
Oy 61 GACCAAAATTTTCATGCCCTTTTAGACGCGATCTTGAACAGATCCGAACACACCCCTT 120
      || || || || || || || || || || || || || || || || || || || || ||
Db 1046 GACCGGATGAGGACACCGCTCTCGATGCCGTGCGCCCGGAGACCGCGGGGCGACG 1105
Oy 121 GCTTCGGAACCTACGTAAACCGGATGCTTTTACTGGCGGGGGAATCATCCACGACG 180
      || || || || || || || || || || || || || || || || || || || || ||

```

```

Db 1106 GCCGTCAGACCCCTGATCACCACCGGTCAAGTCATATCCCGGCGAGCTCACCAACACC 1165
Oy 181 GGCTGGGTAGACATCGAAGAAGATCAACCCCTAACCCGTTCCGGAATTTGGCATGTGCAT 240
Db 1166 GGCTAGCGCCCGCCATCGCCCACTGGTGCGGACACCGCTCTCTTCATCGGATCAGACTCC 1225
Oy 241 TCCGACATGGGCTTGGACCTACTCTGCTGCGGTTCTGAGCGCTATCGGCAAAAGCTCT 300
Db 1226 TCGGCCAAGAGGCTTGCAGCGCCCTCTGTCGGGGGTTCGCTCTCCATCGGCGCGACTCC 1285
Oy 301 CCTGACATCAACCAAGGCGCTTAC----- 324
Db 1286 CCGGACATCGCCCGGGTGGACACCGGTACAGACGGCGGGGCGGGGCAACGGCCCGC 1345
Oy 325 -----CGTCCGATCCCTGTGACAGCGCGCGGCTGACCAGGCTTGATGTTGGC 375
Db 1346 GCGCGACCGGCTGACAGCTGACCGCGCGCGCGCGCGCGCGCGCTGATGTTCGCGC 1405
Oy 376 TAGGCACTAATGAACGACAGCGTGTGATCCAGCAGCATATCACTATGACACCGCTTG 435
Db 1406 TAGGCTGGACGAGACCCCGAGCTGATGCGCTCCGATCACTGGCGCACCGGCTC 1465
Oy 436 GTACACCTGACAGCTGAGTGGCTAAAGAGGCACTCTCCGCTGCGCGCGCGAGCGG 495
Db 1466 TCCCGCGGCTGTGCGAGGTGGGGAAGAACGCGACGATCCCTACCTCCGCGCGAGCGG 1525
Oy 496 AAAAGCAAGTACTTTAGTATGACGAGCGCAAAATGCTGTGATCATGCTGTGCG 555
Db 1526 AAGACCAAGTACCATCAAGTACGAGCGGACAGCGGCTCGCTTCGACAGAGGTG 1585
Oy 556 CTTTCACCTACGACCTCTAAGAGATGACAGAAATCGCTGCAAGACCGGTAATGAA 615
Db 1586 GTCTCTCCAGACGCGCTCCGCAATGACCTGACCTGCTACTGCGCGCGCGACATCCCG 1645
Oy 616 GAGATCATCAAGCAATTTCTGCGGCTGAATGCTGACTTTCGCAACAAATTTTCA-- 673
Db 1646 CCGCATGTGTGAGACCCCTCTCGCGGAGCTGGCGAGACGCGCATCAAGCTGACACCC 1705
Oy 674 -----TCAACCGACCGGCTGTTGCTTATCCGCTGGCCCAATGGGT 714
Db 1706 GCGGCGTACCGGCTGTGTGATCAATCCACCGCGGCTGTGAGATCGGCGCGGATGGC 1765
Oy 715 GACTGCGCTGTGACTGCTGTAATTAATGTTGATACCTACGCGCGCATGGCGCGCTAC 774
Db 1766 GACGCGGCGCTGACCGCGCGCAAGATCATTCGACACTACGCGCGCATGGCGCGCAT 1825
Oy 775 GGTGGCGGTGCAATCTCTGTGAAGATCATCAAAAGTGAACCGCTTCCGAGCTACGCA 834
Db 1826 GCGGCGGTGCTCTCCGGAAGAGACCGCTCCAGAGTGAGCGGCTTCCGCGCGCTACGCA 1885
Oy 835 GCACGTTATGTGCGCAAAACATCTGCTGCTGCGCTGGCGCATGCTTGTGAATTCAG 894
Db 1886 ATCCGCTGTGGTGGCCAAAGACGTGCGGCGCGGCTGGCTCGCGGCTGTGAGTGAG 1945
Oy 895 GTTTCCTAGCAATGCGCGGTGTGAACGACCTCCATCATGTTAGAAATTTCCGTTACT 954
Db 1946 GTTCGCTTACGCTTCCGCAAGCGCGAGCGGTGCTGTGCTGAGACACTTCCGACCC 2005
Oy 955 GAGAAAGCTCTTGAAGAACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1014
Db 2006 GCGACGCTGAGAGCTGAGCGGATCGAGACGAGCATGCGGAGGTTTCGACCTCCGCGC 2065
Oy 1015 TAGGCTGTATTCATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074
Db 2066 GCGCGCATATTCGCGGATCTGAGACCTGCTCGCGCATCTACCAAGACCGCGCTGAC 2125
Oy 1075 GGTCACTTTGCTGCTGAAC-----ATTTCCTGGGAAAAAACCGCAAGCGCGAGCTG 1128
Db 2126 GGTCACTTTGCTGCTGAAC-----ATTTCCTGGGAAAAAACCGCAAGCGCGAGCTG 2185
Oy 1129 CTGCGGATGCTGCGGCTT 1148
Db 2186 CTGATCGCGCGCGCGCGCT 2205

```

```

RESULT 3
US-08-955-957A-6
: Sequence 6, Application US/08955957A
: Patent No. 6312920
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Rostock Jr., Paul R.
: TITLE OF INVENTION: SAM Operon
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: U.S.
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/955,957A
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Webster, Thomas D.
: REGISTRATION NUMBER: 39,872
: REFERENCE/DOCKET NUMBER: P-10162
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3334
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4848 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: mRNA
: HYPOTHEtical: NO
: ANTI-SENSE: NO
: US-08-955-957A-6

Query Match          32.4%  Score 374:  DB 4:  Length 4848:
Best Local Similarity 49.7%:  Pred. No. 4, 2e-101:
Matches 606:  Conservative 122:  Mismatches 420:  Indels 72:  Gaps 3:

Oy 1 ATGGCAAAACACCTTTTACGTCGAGTCCGTTCTGAAGGCGCATCTGACAAATTTGCT 60
Db 986 AUCUCACGUGCGCUCUGUACACUCUGAGUCCUGAGCGAGGCGCACCCGCAAGAUCCG 1045
Oy 61 GACCAAAATTTGTGATGCGCTTTAGACGGGATCTCTGAACAGAGATCCGAAGCACCGTT 120
Db 1046 GACCGGAUACGAGACCGCUCGUCGAGCCUGUCGCGCGGACCGCGCGCGCAAGGUC 1105
Oy 121 GCTTGGAAACCTACTATAAAACCGGATGTTTGTGGGCGCAAAATTCACCAACCCAGC 180
Db 1106 GCGGUGGACACCCUGAUACACACCGGUGCAAGUCUUAUUGCGCGGCGCAAGUACACAC 1165
Oy 181 GCCTGGGTAGATCGAAGAAGATCAACCCGTTTCGGAATTTGGCTATGTGCAT 240
Db 1166 GCGUACGCGCCCAUCCGCCCAACUGUGGCGGACACCGGUCUUGCCAGUGUACGACUC 1225
Oy 241 TCCGACATGGGCTTTGACGCTTAATCTCTGT3CGGTTTCTGAGCGCTTATCGGCAAAAGCT 300
Db 1226 UCGGCCAAAGGCUUACGCGCGCUCGUCGCGGUGUCGUGUCCAUUCGCGCGCAGUCC 1285
Oy 301 CCTGACATCAACCAAGGCGCTTAC----- 324
Db 1286 CCGGACAUCCGCGGUGUGGACACCGGUAACGACGCGCGGCGGCGGCGGCGGCGGCGG 1345
Oy 325 -----CGTCCGATCCGCTGGAACAGACGCGCGGTGACAGAGCTGTGATGTTGGC 375

```



```

0Y 301 CTTGACATCAACAGGCGCGTTGACGCTGG-----C 330
Db 301 CCGGACATGCCCCAGGGCGCTGCACAGCCCTTACAGAACCCGCGTTGAGGCGGAGGACAC 360
0Y 331 GATCCGCTGGAAAGGGCGCGGGGTGAACAGGGTCTGATGTTGGTACCAACTAATGA 390
Db 361 GAGCTGACCAAGAGGGCGCGCGGCACACAGGGCGCTGATGTTGGCTAGGCCACGAGAG 420
0Y 391 ACCGACGTGCTATGCCAGCACCTATCACCTATGCACACCGCTCTGTGTACAGCTAGGCT 450
Db 421 ACCCGCTCGCTATGCCGTGGCCCATCGAGCTGCGCCACCGCCTCTCGCGCGGCTCAC 480
0Y 451 GAAGTGGCTAAAAACGGCACTGCGCGGTGGCTTCCCGCCGACGGCAAAAGCAGTGTACT 510
Db 481 GAGGTCCGCAAGAGGACGCGACCGTGTCCCTTACGTGCGCCCGCAGGCAACACCGAGTCA 540
0Y 511 TTTCAGTATGACAGCGGCAAAATCGTTGGTATGATGCTGTGCTGTTTCCACTAGCAC 570
Db 541 ATGAGATACCAAGGAGGACGCGCGCGCGGCTGTGAGAACCGTGTGCTCTCTCCACGAC 600
0Y 571 TCTGAACAGATCGACCAAGAAATGCGTGCAGAAGACGGGTAAATGAAGATCAAGCCA 630
Db 601 GCGCGCCACATCGACCTCGGCTCCGTCTGCTCACCCCGACATCCGACACAGTGTGTGAG 660
0Y 631 ATTTCGCCCCCTGAATGCTGACATTTGTGGCACCAGATTC----- 670
Db 661 CACGTCTCTCCCGCCTACGCGCGAGGACGGCATCAATGCTGAGACGAGCACTACCGCTG 720
0Y 671 -TCATCAACCGGACCGGTGCTTTCGTTATCGGTGGGCGCAATGGGTAGCTGCGGTGACT 729
Db 721 CTGGTCAACCGGACCGGCGGCTTTTCGAGATCGGCGGCGCGATGGCGGACCGGCTGACC 780
0Y 730 GGTGCTAAATTTATCGTTGATATCATCGGCGGCGATGGCGGCTCACGGTGGCGGTGCATTC 789
Db 781 GCGCGCAAGATCATCATCGACAGTACGGCGGCGATGGCGCGCACGCGGGTGGCGGCTTC 840
0Y 790 TCTGCTAAGATCATCAAAAGTGAACGCTTCCGCAAGCTACGCAACACTTATGTGGCG 849
Db 841 TCCGCGCAAGGACCCGCTCAAGGATGCAACCGTTCGCGCGCTACGCGATCGCGTGGGTCC 900
0Y 850 AAAACATCGTTGCTGCTGCGTGGCGGCGGAGTCTGTTGAAATTCAGATTTCCACGCAATC 909
Db 901 AAGACGCTCTTCCCGCGGCGGCGCTCTGCGCTCTCCGCTGCGAGGTCCAAGTGCCTAC 960
0Y 910 GCGCTGGCTGAACGCACTTCATCATGATGAGAATTTTCGTACTAGAGAAAGTGCCTTCT 969
Db 961 GCGCAAGGCGGACGCGGCGGCGGCTGTTGCGTGAAGAGCTTGCGGACCGGCGCGTCC 1020
0Y 970 GAACACACTGACCTGCTGCTGATGCTGATGCTTTCGACCTGCGCCCATACGGTCTGATTAG 1029
Db 1021 GAGGCGCATTCAGAAAGGCAATCACGCGAGGTCTTTCGACTGCGCGCGCGCTCATTCG 1080
0Y 1030 ATGCTGATGATGCTGACACCGGATCTAACAAGAAACCGGACGATACGGGTACATTTGGT 1089
Db 1081 GACCTGACACTGCTGGGCGCCATCTACGCGCGCACCGCGCTACGGGCACCTTGCGCGC 1140
0Y 1090 GAAC-----ATTTCGCGTGGGAAAAAACCGCAAGGCGCAGCTGCTGCGGATCTGCC 1143
Db 1141 GAACATGCGGACACTTCACTGATGGAGGGGAGACGACCGCGCCACGCGCTCAAGGCGCG 1200
0Y 1144 GGTCT 1148
Db 1201 GGTCT 1205

RESULT 5
US-08-510-646B-4
: Sequence 4, Application US/08510646B
: Patent No. 6077699
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Bianche, Francis
: APPLICANT: Crouzet, Joel

```

```

APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibault, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flunegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/C0923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208
US-08-510-646B-4

Query Match 32.0%; Score 369.8; DB 3; Length 1208;
Best Local Similarity 59.4%; Pred. No. 4e-100;
Matches 716; Conservative 0; Mismatches 432; Indels 57; Gaps

QY 1 ATGCCAAACACGCTTTTACGTCGCAGTCGCTCTGTGAAGGCGCATCTGACAAATTCCT 60
Db 1 ATGTCGCCGTCGCTCTTACCTCGSAGTCCTGTGACCGAGGGCCACCCGACAAATTCGCC 60
QY 61 GACCAAAATTTCTGATCCGCTTTTAGACGCGCATCTCTGAAACAGATCCGAACGACGCGTT 120
Db 61 GACCAAGATCACTGACACCGCTCTGACGCGCTGCTGCGGAGAGACCCGCGCTACGCGTC 120
QY 121 GCTTGGCAAACTCTAGTAAAAACGGGCACTGGTTTATGTTGGGGGGAATATCAACACAG 180
Db 121 GCGGTGCGAGACCTGATACACACGGGCGGCAAGTCCACATCGCGGAGAGTCAACACAG 180
QY 121 GCGGTGCGAGACCTGATACACACGGGCGGCAAGTCCACATCGCGGAGAGTCAACACAG 180

```



```

OY 181 GCCTGGTAGACATCGAAGATACCCGTAACCCGTTGCGGAATTGGCTATGTGCAT 240
DB 181 GCGTACCGCGCCATCGCCCACTGGTGTGGGACAGATCCTGGCGCTAGGACTGCG 240
OY 241 TCCGACATGGCGTTGACGCTTAACCTGCTGCGGTTGTGAAGCCGTATCGCAACAGCTCT 300
DB 241 TCCGCGCAAGGGCTTGCAGAGCGCGCTCTCGCGGCTCTCCGTTCTCATCGGCGCCAGTCC 300
OY 301 CCTGACATCAACGAGGCGTTGACCGTGC-----C 330
DB 301 CCGGACATCGCGCCAGGCGCTGACAGCGCCTACAGAGACCGCGCTGAGGCGAGAGAGAC 360
OY 331 GATCCGCTGGAACAGGCGCGGCTGACAGGCTGTGATGTTTGGCTACGCACTAATGAA 390
DB 361 GAGCTCAGCAGCAGGCGCGCGCGGCTGATGTTGCGCTACGCGCAGCAGAGAG 420
OY 391 ACCGACGTGCTGATGCGCAGCAGCTATCAGCTACACAGCGTGTGATACAGCGTCAAGCT 450
DB 421 ACCCGCTCGGTGATGCGCGCTGCGCTATGAGCTGCGCCACCGCTCTCGCGCGCTGACC 480
OY 451 GAACTGGCTAATAAAGGCGCTCTGCGTGGCTGCGCGCGAGAGCGGAAAGCCAGCTGACT 510
DB 481 GAGGTCCGCAAGGAGCGGCGCGCTGCGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 540
OY 511 TTTCAGTATGACGAGCGGCAAAATCGTGTGATGATGCTGTGCTGCTTTCACACTAGACAC 570
DB 541 ATTCAGTATGACGAGCGGCGCGCGCTGCGCTGCGCGCGAGAGAGAGAGAGAGAGAGAG 600
OY 571 TCTGAAGATGACGAGAGAAATCGCTGCAAGAAACGGTAAATGGAAGATATCATCAACCA 630
DB 601 GCGCGCGCATCATGACATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
OY 631 ATTCTGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
DB 661 CACCTCTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
OY 671 -TCATCAACCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
DB 721 CTGCTCAACCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
OY 730 GCTGCTAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 789
DB 781 GCGCGCAAGATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 840
OY 790 TCTGCTAAGATCATCAAAAGTGAACCGTTCCGACCTTACGACGACGACGACGACGACG 849
DB 841 TCCGCGCAAGGAGCGGCTGCAAGGCTGACGCTGCGCGCTGACGAGATGCGCTGCTGCTG 900
OY 850 AAAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909
DB 901 AAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
OY 910 GCGCTGCTGAACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
DB 961 GCGAAGGCGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
OY 970 GAACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1029
DB 1021 GAGCGCATGAGAGGCGCTATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
OY 1030 ATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
DB 1081 GACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
OY 1090 GAAC-----ATTCCCGTGGCAAAAAAGCAGAAAGCGAGCTGCGCGGATGCTGCGC 1143
DB 1141 GAACATGCGCGACCTTCACTGTGGAGACGAGACCGACCGCGCCACCGCGCTCAAGCGCGCG 1200
OY 1144 GGTCT 1148
DB 1201 GGTCT 1205

```

```

RESULT 6
US-09-231-818-4
: Sequence 4, Application US/09231818
: Patent No. 6171846
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanc, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: TITLE OF INVENTION: Polypeptides Involved in the
: TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleoside Sequences
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/231,818
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,852
: FILING DATE: 10-MAY-1995
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/11441
: FILING DATE: 25-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806, 0054-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ. ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1208 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: S.pristinaespiralis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1208
: US-09-231-818-4
:
: Query Match 32.0%; Score 369.8; DB 4; Length 1208;
: Best Local Similarity 59.4%; Pred. No. 4e-100;
: Matches 716; Conservative 0; Mismatches 432; Indels 57; Gaps 3;
OY 1 ATGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGCGCATCTGACAAATTCCT 60
DB 1 ATGTCCCGTGGCTGCTTACCTCGAGATCGGTGACCGAGGCGCACCGCAAGATCGCC 60

```

```

Db      1141 GAACTGCCCGCCACTTGTACCTGGAGCAGGACCACCGCGCGCCACCGAGCTCAAGGCCCGCGCC 1200
QY      1144 GGCTCT 1148          11111
        1201 GGCTCT 1205          11111

RESULT 7
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanchet, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibault, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
Prior APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
Prior APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.plistinaespiralis
US-08-403-852D-1

Query Match 32.0%; Score 369.8; DB 2; Length 5392;
Best Local Similarity 59.4%; Pred. Mismatches 100; Indels 57; Gaps 3
Matches 716; Conservative 0; Nonmatches 432;
1 ATGCACAAAACACCTTTTACCTGCAGTCGCGTCTCTGAAGGCGCATCCGTACAATAATTCCT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Query Match 32.0%, Score 369.8; DB 3; Length 5392;
 Best Local Similarity 59.4%; Pred. No. 7.8e-100;
 Matches 716; Conservative 0; Mismatches 432; Indels 57; Gaps 3;

OY 1 ATGGGAACACCTTTTACGTCGAGTCGCTCTGAAAGGCACTCCGACAAATTCGT 60
 DB 3558 ATGTCCTCCCTCCCTTACCTTCGAGTCGAGACCGGACCCGCAAGATCCGC 3617
 OY 61 GACCAAAATTTTGATGCGCTTTTACGCGATCTCGAAGAGATCCGAAAGACAGCGTT 120
 DB 3618 GACCAAGATCACTGACACCGCTCTCGACGCTGCTGCGGAGAGACCCGCTCAGCGCTC 3677
 OY 121 GCTTGCGAAACCTTACGTAACAAACCGGATGTTTGTAGTGGCGGAAATGACACAGC 180
 DB 3678 GCGGTGAGACCTTGATCAGACCGGCGGAGTCACATCGCGGAGAGTGCACACAG 3737
 OY 181 GCGTGGGTAGACATCGAAGAGATCACCCGTACACCGCTTCCGGAATTGGCTATGCT 240
 DB 3738 GCGTACCGCCCATCTGCGCACTGCTCGGACACGATCTGGCCATCGCGTACGACTCG 3797
 OY 241 TCCGACATGGGCTTTGAGCTAAGTCTGTGGGTTTGAGGGTATCGGCAACAGTCT 300
 DB 3798 TCCGCAAGGGCTTTGAGCGGCGCTCTGCGGCTGCTGCTCCATCGCGCGAGTTC 3857
 OY 301 CCTGACATCAACACAGGCGCTTGACCGTGC-----C 330
 DB 3858 CCGGACATCGGCGGACGAGGCTCGACAGCGCTACGAGACCGCGTGCAGGCGAGCAGC 3917
 OY 331 CATCGCGTGAACAGGCGGCGGTGACAGGCTGTGATGTTTGGCTACGCACTAATGAA 390
 DB 3918 GAGCTCGACAGCAGGCGGCGGCGGACGAGCGCTGATGTTTGGCTACGCGCCACGACAG 3977
 OY 391 ACCGAGCTGTGATGCTCCAGACCTATGACACACCGTCTGTAGAGGTGAGGT 450
 DB 3978 ACCGCTCTGCTGATGCTCCCTGCTCCATGACGCTCGCCACCGCTTCCGCGCGGTCAAC 4037
 OY 451 GAAGTGCCTAAACAGGCACTTGCCTGCGTCCCGCGGACCGCAAAAGCAGGTGACT 510
 DB 4038 GAGGTCCGCAAGAGAGGAGCAGCGTCCCTACGTCGCGCGCGGCAAGCAGAGTCAAC 4097
 OY 511 TTTGAGTATGACAGGCGCAAAATCTTGTATGATGCTGTGCTTTCCTACAGCAG 570
 DB 4098 ATCGGTACAGGCGGAGCGCGCGGCTGCGCGTGCAGACCGCTGCTGCTCCCGACAG 4157
 OY 571 TCTGAAGAGATGACGCAAAATCGCTGCAAGAAGAGGTAATGGAAGATCATCAAGCA 630
 DB 4158 GCGCGCGGACATGACCTGCGCTCCTGCTCACCGCGGACATCGCGGAGACTGCTCAG 4217
 OY 631 ATTCTGCCGCTGAATGCTGACTTCTGCAACCAATTC----- 670
 DB 4218 CACGTCTCGCGCGGACGCGGAGAGGATCAAGCTCGAAGAGGACACTACCGCGCTG 4277
 OY 671 -TCATCAACCCGACCGGCTGCTTTCGTTATCGTGGCCCAATGGGTGACTGCGCTGACT 729
 DB 4278 CTGCTCAACCCGACCGGCTGCTTTCGATGATCGCGCGGATGCGCGGCGCTGAGC 4337
 OY 729 GGTCTAAATTAATGCTGATACCTAAGGCGGATGCGCGCTACGAGTGTATGTCGCG 849
 DB 4338 GCGCGCAAGATCATATGACACGATGCGGCGGATGCGCGGCGGAGTGGCGGCTTC 4397
 OY 790 TCTGTAAAGATTCATCAAAAGTGAGCGCTTCCGAGGCTACGACGAGCTTATGTCGCG 849
 DB 4398 TCCGCGAAGAGACCGCTGCAAGGTGACCGCTTCCGCGGATGCGCGGCTGCGC 4457
 OY 850 AAAAAATCGCTTGTGCTGCGCTGCGCATGCTTGAATTCAGGTTTCTCAGCAATC 909
 DB 4458 AAGAAGCTGCTGCGCGCGGCTCCCTCGCTCGCGGAGGTCGAGGCTGCGCTACCCCATC 4517
 OY 910 GCGGTGCTGAACGACCTCATCATGTAGAAATTTGCTAGTACGAGAAATGCTCTCT 969
 DB 4518 GCGAAGCGCGAGCGGCTGCTGCTGCTGAGAGCTTTCGCGGAGCGGACGCTGCCAG 4577
 OY 970 CAACAACCTGACCCCTGCTGATGATGTTTTCGACCTGCGCGCATACGCTGATGAG 1029

DB 4578 GAGCGCATCGAAGAGCCCATACCGAGGCTTTCGACCTGCGCGCGCGCATCATCCGC 4637
 OY 1030 ATGCTGATCTGCTGACCCGCTATCAAGAAACCGAGCATACGCTACTTGGTCT 1089
 DB 4638 GACCTGACCTGCTGCGCGCATCTACCGCGCACCGCGGCTTACGGGCACTTCCGCGC 4697
 OY 1090 GAAC-----ATTTCGCGTGAAGAAACCGACAAACCGGACGCTGCTGCGGATCTGCC 1143
 DB 4698 GACCTGCGCGGACTTCACTTGGAGGAGACGACCGCGCGCATCAAGCGCGCGC 4757
 OY 1144 GCTCT 1148
 DB 4758 GCTCT 4762

RESULT 9
 US-09-231-818-1
 Sequence 1, Application US/09231818
 Patent No. 6171846
 GENERAL INFORMATION:
 APPLICANT: Blanc, Veronique
 APPLICANT: Blanche, Francis
 APPLICANT: Crouzet, Joel
 APPLICANT: Jacques, Nathalie
 APPLICANT: Lacroix, Patricia
 APPLICANT: Thibaut, Denis
 APPLICANT: Zagorec, Monique
 APPLICANT: Debussche, Laurent
 APPLICANT: De Grey-Lagard, Valerie
 TITLE OF INVENTION: Polypeptides Involved in the
 TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleotide Sequences
 TITLE OF INVENTION: Coding for These Polypeptides And Their Use
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/231,818
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/403,852
 FILING DATE: 10-MAY-1995
 APPLICATION NUMBER: PCT/FR 93/00923
 FILING DATE: 25-SEP-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/11441
 FILING DATE: 25-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meyers, Kenneth J.
 REGISTRATION NUMBER: 25,146
 REFERENCE/DOCKET NUMBER: 03806.0054-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5392 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO

```

: ORIGINAL SOURCE:
: ORGANISM: S.pristinaespiralis
: US-09-231-818-1

Query Match      32.0%; Score 369.8; DB 4: Length 5392;
Best Local Similarity 59.4%; Pred. No. 7.8e-100;
Matches 716; Conservative 0; Mismatches 432; Indels 57; Gaps 3:

Oy 1 ATGGCAAAACACCTTTTACGTCGAGTCCGTCCTGTAAGAGGATCCTGTGAATAATTGCT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3558 ATGTCCTCGCTGCTGTCACCTCGAGATCCGTACCGAGGCGCCACCCGACAAAGATGGC 3617
Oy 61 GACCAAAATTTCTGATGCGCTTTTACGACGATCCTCGAACAGATCCGAAGACGGCTT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3618 GACCAATCTGATGACACGCTGCTGAGCGCTCTGCGGACGACCCGCTCAGCGCTC 3677
Oy 121 GCTTGGAAACCTACGTAACAAACCGGATTTTACTGGGGGCAATACCAACAGC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3678 GCGGTGAGACCTGATACACCGGCGGATCCACATCCCGGCGAGTCCACCAAG 3737
Oy 181 GCTGCGTAGACATGAGAGATCACCCGTACACCGCTTGGCAATTTGCTATGTGCAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3738 GCGTACCGGCGCATCCGCAATGCTGCGGACAGATCCTGGCGCATCGGCTACGACTCG 3797
Oy 241 TCGGACATGGGCTTTGACGCTACTCTGTGCGGTTCTGAGCGCTATCGGCAACACTCT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3798 TCGGCAAGGCTTTCAGCGGCGCTCTGCGGCTCTCGTTCATCGGCGCGCACTGC 3857
Oy 301 CCTGACATCAACGAGGCGGCTTTCAGCGCTG-----C 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3858 CCGGACATCGGCGCGGCGCTGACAGCGCTACGAGACCGCGCTCGAGCGGCGAGAGAC 3917
Oy 331 GATCGCTGAAAGGCGGCGGCTGACAGGCTGTGATGTTTGGCTACGCACTAATGAA 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3918 GAGCTGACAGCAGGCGCGCGGCGGACGAGCGCTGTGTTGCGCTACGCGACGAGAG 3977
Oy 391 ACCGACGCTGATGCCAGCACTATACCTATGACACCGCTCTGTGTAACGCTGAGCT 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3978 ACCCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4037
Oy 451 GAATGCTGTAAGAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4038 GAGTCCGCAAGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4097
Oy 513 TTTGAGTATGACGAGCGGCAAAATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4098 ATCGAGTACAGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4157
Oy 571 TCTGAAGATGACGACGAATAATGCTGCAAGAACCGTAATGCAAGATGATCAAGCA 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4158 GCGCGCGACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4217
Oy 631 ATTGCGCGGCTGATGAGCTGACTTCTGCGCACAAATTTCT----- 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4218 CAGCTCTCGCGCGCATGCGCGAGAGCGCATCAAGCTCGAGAGCAACTACCGCTG 4277
Oy 671 -TCATCAACCGGACCGGCTGTTTCTGTTATGCGTGCGCCCAATGCTGCTGCTGCTGCT 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4278 CTGCTCAACCGGCGGCGGCTTTCGAGATCGCGCGCGGCGGCGGCGGCTGACG 4337
Oy 730 GGTGCTAAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4338 GCGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4397
Oy 790 TCTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4398 TCGCGCAAGAGCGGCTGCAAGGTCGACGCTTCCGCGGCTGACGCGATCGCTGCTGCT 4457
Oy 850 AAAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4458 AAGAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4517
Oy 910 GCGGTGCTGAAAGCGACTCATGATGATGATGATGATGATGATGATGATGATGATGAT 969

```

```

Db 4518 GCGAAGCGCGAGCGGCTGCGCTGCTGCTGAGACGTTGGGACCGGACCGCGCCAG 4577
Oy 970 GAACAACGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4578 GAGCGCATGAGAGGCGCATACACGAGGCTTGTGACCTGCGCGCGCGCGCATATCCG 4637
Oy 1030 ATGCTGATGCTGCTGACCGCATCTACAAAGAACCGCAGCATACGCTACCTTGTCT 1089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4638 GACCTGACCTGCTGCGCGCATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4697
Oy 1090 GAGC-----ATTCCCGTGGGAAAGAACCGAAGCGCAGCTGCTGCGCGATGCTG 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4698 GAACGCGCGGCTTCACTGAGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4757
Oy 1144 GGTCT 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4758 GGTCT 4762

```

```

RESULT 10
US-09-273-686-1
: Sequence 1, Application US/09273686
: Patent No. 6228625
: GENERAL INFORMATION:
: APPLICANT: Zalcatal, Magdalena
: APPLICANT: Burnham, Martin K. R.
: APPLICANT: Bismas, Sanjoy
: APPLICANT: Brown, James
: APPLICANT: Ingraham, Karen, A.
: APPLICANT: Chalker, Allison F.
: APPLICANT: So, Chi Y.
: APPLICANT: Holmes, David J.
: APPLICANT: Van Horn, Stephanie
: APPLICANT: Warren, Richard L.
: TITLE OF INVENTION: mek
: FILE REFERENCE: GM10176
: CURRENT APPLICATION NUMBER: US/09/273,686
: CURRENT FILING DATE: 1999-03-22
: EARLIER APPLICATION NUMBER: 60/106,767
: EARLIER FILING DATE: 1998-11-03
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1191
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: US-09-273-686-1

```

```

Query Match      31.9%; Score 368.2; DB 4: Length 1191;
Best Local Similarity 60.3%; Pred. No. 1.2e-99;
Matches 676; Conservative 0; Mismatches 408; Indels 33; Gaps 3:

Oy 14 TTTTACGTCGAGTCTGCTGCTGTAAGGCGATCCTGACAAATTTGCTGACCAATTTCTG 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 TATTCAGCTCTGATCTGATCTGATCTGATGAGGCGATCGGATTAAGATTGACAGCAATTT 76
Oy 74 ATGCGGTTTATGACGCGCATCTCTGACAGAGATCGGAAGACGCGCTTTCGGAACCT 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 ATGAGATTTGAGATGATATTTTACAAAGATCCAGAGGCGCATGCTGCTGCTGCTGCT 136
Oy 134 ACGTAAGACCGGATGTTTATGTTGCGGCGGAATTCACACGAGCGCTGGGTAAACA 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 CTGATATATACGCTTCTGCTCAGCTTTTGTGAATTTCTACAAATGCGCTATGCTGATA 196
Oy 194 TCGAAGATACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 TTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
Oy 254 TTGACGTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 TTTCTGCTGAGACGCTGAGATACACCACTTTGTTGTAACAAATCTCTGACATCTGCT 316

```

```

OY 314 AGGCGTTGACCGTGC-----GATCCGCTGGAACAG- 345
DB 317 AAGGTGTTAAGACAGCCTTGAGGTTGCGGAATGCTGATCAGATCCATCGAGCTTGA 376
OY 346 --GGCGCGGTACCAAGGCTCTGATGTTGGCTACGCAACTAATGAACCGCAGCTGA 403
DB 377 TTGGAGCAGGTGACCAAGGCTCATCTTTGGATTGGCATAGATGAAGAGAGCTTAA 436
OY 404 TGCCAGCACTTCACTTGTGACACCGCTGCTAGCAGGCTGAGAGCGTGAANA 463
DB 437 TGCCATTGCTCAATTCAGTCACTGATCAATTAATTCGCTGCTGAGAGACTTCGAACT 496
OY 464 AGGCACTCTGCGTGTGCTGCGCCGAGCGCAAAAGCCAGGTGACTTTTCAGTATGAG 523
DB 497 CTGGAGAATTAAGTATCTCCGCTCCAGATGCAAAATCACAACTTACAGTTAGTACGAG 556
OY 524 AC--GGCAAAATCGTTGGTATGATGATGCTGCTGCTTTCCACTCAGCAGCTGAAGA 580
DB 557 AAAATGACCGTCCGGTACGTAGATACAGTGTATTCTACTGAGCATGATCCAGAGG 616
OY 581 TCGACCAAGAAATCGCTGCAAGAGCGGTAAATGGAAGATCATCAAGCAATTCGCCGG 640
DB 617 CCAGTAATGAACAAATCCATCAAGATGTGATTGACAGAGGTATCAAGAGTTATTCAT 676
OY 641 CTGAATGCTGACTTCTGCGCAAAATTCCTCATCAACCGCAGCGTCTGTTATGATCG 700
DB 677 CTCTTATCTGATGATAGCAAAATTCCTTATCATGCGCAGAGCTGCTTTGTTATG 736
OY 701 GTGCGCCAAATGGGTAGCTGCGGTCTGACTGTGCTGTAATAATTCGTTACCTACGCGC 760
DB 737 GTGTGCTCAAGGGAGCTGAGGTTGACGTGCTGTAAGATTATGATGATACCTTATGCTG 796
OY 761 GGTATGCGGTGACGCTGCGGTGCTGCTGTAATAATTCATCAAGAGGAGGAGCGCT 820
DB 797 GCTACTCTGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856
OY 821 CCGCAGCCTTACCCAGCAGCTTATGCTGCGCAAAAGCACTTCTGCTGCTGCTGCTGCTGCT 880
DB 857 CAGCCTTATGCGGCTGCTGCTTATGCTGCGCAAAATTCGTTGACAGAGCGCTTGGTGA 916
OY 881 GTTGTGAATTAAGCTTCTGACCAATCGCGGTGCTGACCGACCTCATATGCTGAG 940
DB 917 AGCGCAAGTAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976
OY 941 AACTTGTGCTAGTGAAGAGGCTTGTGAACACTGACCTGCTGCTGCTGCTGCTGCTGCT 1000
DB 977 AATCTTGTGCTAGTGAAGAGGCTTGTGAAGAGGCTTGTGAAGAGGCTGCTGCTGCTGCT 1036
OY 1001 TCGACCTGCGCCATACGCTGCTGATTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1060
DB 1037 TTGACCTTGGCGCTGAGGATTAATGCAAAATGCTGAGCTCAAGCTCAATTTACCGTC 1096
OY 1061 AAACCGACAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
DB 1097 AAACATCGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133

```

RESULT 11
US-08-961-527-163
Sequence 163, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:
APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland

COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 163:
SEQUENCE CHARACTERISTICS:
LENGTH: 8494 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-163

Query Match      31.9%      Score 368.2      DB 4      Length 8494:
Best Local Similarity 60.5%      Pred. No. 2.9e-99:
Matches 676: Conservative 0: Mismatches 408: Indels 33: Gaps 3:

OY 14 TTTTACGTCGAGTCGCTCTCTGTAAGGCGATCTGACAAATTCCTGACCAATTTCTG 73
DB 4736 TATTCACGCTGTAATCTGATCTAGAGGCGCATCCGATTAAGTTGACAGCAAAATTTGCG 4795
OY 74 ATGCGCTTTTACAGCGATCTCTGCAACAGATCCGAAGCAGCGCTGCTGCGAACT 133
DB 4796 ATGCGATTTTGGATCTTATTTAGCAAGGATCCAGAGCGCGCTGCTGCTGCGAAACG 4855
OY 134 ACGTAAACCGGCAATGCTTATGTTGGCGGCAATCAACACAGCGCTGCTGCTGCTGCTGCT 193
DB 4856 CTGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4915
OY 194 TCGAAGAGATCAACCGTACACCGCTTCCCAATTTGCTATGCTGCTGCTGCTGCTGCTGCT 253
DB 4916 TTAACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4975
OY 254 TTGACGCTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
DB 4976 TTTCTGCTGAGAGCGGTGAGGATCAACCCATCTTTGCTGGAACAATCTCTGACATCGCTC 5035
OY 314 AGGCGGTGACCGTGC-----GATCCGCTGGAACAG- 345
DB 5036 AAGGTGTTAAGCAGCGCTTGGAGGTTCTGSAATTCGATCAAGATCCATCGAGCTTGA 5095
OY 346 --GGCGCGGTACCAAGGCTCTGATGTTGGCTACGCAACTAATGAACCGCAGCTGCTGA 403
DB 5096 TTGGAGCAGGTGACCAAGGCTCATCTTTGGATTGGCATAGATGAAGAGAGAGCTTAA 5155
OY 404 TGCCAGCACTTACACTATGACACACCGTCTGCTGACAGCGTGAAGTGAAGCTGAANA 463
DB 5156 TGCCATTGCTCAATTCAGTCACTGATCAATTAATTCGCTGCTGCTGCTGCTGCTGCTGCT 5215
OY 464 AGGCACTCTGCGTGTGCTGCGCCGAGCGCAAAAGCCAGGTGACTTTTCAGTATGAG 523
DB 5216 CTGGAGAATTAAGTATCTCCGCTCCAGATGCAAAATCACAACTTACAGTTAGTACGAGT 5275
OY 524 AC--GGCAAAATCGTTGGTATGATGATGCTGCTGCTTTCCACTCAGCAGCTGGAAGA 580
DB 5276 AAAATGACCGTCCGGTACGTAGATACAGTGTATTCTACTGAGCATGATCCAGAGG 5335
OY 581 TCGACCAAGAAATCGCTGCAAGAGCGGTAAATGGAAGATCATCAAGCAATTCCTGCCCG 640
DB 5336 CCAGTAATGAACAAATTCATCAAGATGTGATTGACAGAGGTATCAAGAGATTATTCAT 5395

```

Oy	611	CTGAATGGCGACGTCCTGCCACCAAATTCTTCAATACCCGAGCGGTTCGTTATCG	700
Db	5396	CTTCTTAATCTTGATGATAMAGCAAATAATCTTTATCATCCTGCAGAGTCGTTTTGAATCG	54555
Oy	701	GTCGCCAATAGGGTGAAGCTGGGTCTGACACTGTCGTAAAATTATCTTATACCTAACGGCG	760
Db	5456	GTGGCTCTCAAGGGAGACAGCTTTGACHTGTCGTAACTTTATTGTACATCTTATGGTG	5515
Oy	761	GCATGGCGCGTCACAGGTGGCGGTGCATTTCTTGTGTAAAGATCCATCAAAAGTGGACGTT	820
Db	5516	GTATCTTCTCTCATGGTGTGGGTGCCCTTCTCTGTAAAAATCGCACTAAAGGTGATCGTT	5575
Oy	821	CCGAGCGCTACGAGACAGCTTATGTGCGCAAAAACACTCTTGCTGAGCGCGCCGATC	880
Db	5576	CAGCTCTTATGCGGCTCGGTATATTGGCCAGAATATCTGTTACAGACACCTTGCTAAGA	56355
Oy	881	GTTGTGAATTCAAGTTCCTTACGACAATCGCGCTGGCTGAAACCGACCTTCATCATGTAG	940
Db	5636	AGGCAGAAAGTGCAGTTGCGCTTATGCTATGCGGTGTCGCCAACCTTTCTGTTCGTATCG	56955
Oy	941	AACTTCGCGTACTAGAAAGTGCCTTCTGAAACAAGTCACTGACCCCTGGTAGAGTGATCT	10000
Db	5696	ATATCTTCGGTACAGACACGTAAGCTTGAAGATCAACTTGTAAAAACGGCTGCTCAATCT	57555
Oy	1001	TGCAGCTGCGCCATACGCGTCTGATTTCAAGATCTGGATCTGCTGACCCGATCTACAAG	10600
Db	5756	TTGACCTTCGCGCTGAGAGGATATATCCAATAGCTGAGACTCAAGCGTCCAATTTACGCTC	5815
Oy	1061	AANCCGAGCATACGGTCACTTTGGTCGTGAACATTT	1097
Db	5816	AAACATCGGCTTACGGTCACATGGGACGATGAGATAT	5852

```

RESULT 12
US-09-134-001C-2240
: Sequence 2240, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucelle-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 2240
: LENGTH: 1248
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-2240

```

Query Match	30.2%;	Score 349.2;	DB 4;	Length 1248;
Best Local Similarity	58.5%;	Pred. No. 5.3e-94;		
Matches 687; Conservative	0;	Mismatches 448;	Indels 39;	Gaps 3

Seq	Accession	Gene	Species	Length
0y	7	AAACACCTTTTACGTCGCGATCCGCTCTCTGAAAGCGATCTCTGCACAAATTTGCTGACCA	66	
	1111			
	1111			
Db	61	AAACGATATTATTACCTCAGAAATCAGTACTAGCAAGCCATCTGTAAATTTGCTGACCAA	120	
	1111			
0y	67	ATTCTGATGCGCGTTTACGAGCGCATCTCTGAAACAGATCTCGAAGCAGCGTTCGCTGC	126	
	1111			
	1111			
Db	121	GTTTCACATGCCATTCTTGATGAAATTTAAAGCATGATCCCAATGCTCTGTAGCATGT	180	
	1111			
	1111			
0y	127	GAACACTACGTAACAAACCGCGATGGTTTATAGTGGCGGCAAAATCACACACGCGCTGG	186	
	1111			
	1111			
Db	181	GAACACACAGTAAACACAGGAATGCGACTTATTTCTGGAAGAAATCTCAACACACTACTTAT	240	
	1111			
	1111			
0y	187	GTAAGACTCAAGAGATCAACCCGTAACACCGCTTTCGGAATTTGGCTATGTGCATTCCGAC	246	
	1111			
	1111			

Db	241	GTGATATTCTCTAAAGTGGTGAAGAAAGCATTTAAAGATATATGTTACTACTAGACAA	300
Qy	247	ATGGGCTTTACCCCTAATCTCTGTGGGTTCTGAGGCTATGCGCAACAGTCTCTGAC	306
Db	301	TATGGTTATATTTACAAACAAATAGCACTGTAACTGGCATTGATGAACAGACTCTCTAT	360
Qy	307	ATCAACCAAGGGCGGTTGACCGCTGC-----	366
Db	361	ATTTCACACAGGCTTAGCAACAAAGCTCTTGAGATATCGAAATGATATTTCTCGAAGACAA	420
Qy	337	CTGGAACAGGCGCGGGGTGACAGGGCTGCATGTTTGGCTACGCACTAATCAAAACCGAC	396
Db	421	GAACCAACAGCTGACAGTGCACCAAGATTAATGTTGGTATAGCACTGATGAACCTGAT	480
Qy	397	GTCGTGATGCCAGCACCCTATACCACTATGCACACCGTCTGGTACACCGTCAGCTGAATG	456
Db	481	ACGATATATGCTCTACTATATTTCTGTGCAATCACTTGTGAACGATTTGGCTGATATGA	540
Qy	457	CGTAAACAGGCACTCTGCCGTGGCTGCGCCCGGAGCGCAAAAGCAGCGTGCATTTGAG	516
Db	541	CGAATAATGAATTTTATAGTATTCTTCGTCCAGATGGAAGATACAGTACTGTGA	600
Qy	517	TA---TGACAGCGCAAAATCGTTGGTATTCATGCTGTCTGCTCTTCCACTACAGCACT	573
Db	601	TATGTTAAATGACAAACCTAGACCTATTGATACCATTTGTTGTTTACACAAACATCTCT	660
Qy	574	GAAGAGATGACCAAGAAATCGTGCANAGACGGGTATATGAAGACATATCAACCAATTT	633
Db	661	GAAATGTTAAGTTAGCACAAATTTGAAAAGACATTTAAACCGCATTTATTACCACACT	720
Qy	634	CTGCGCCGTGAATGCTGACTTCTGCGCACAAATTTCTTCATCAACCGACCGGTCTTTTC	693
Db	721	GTACATTAACCTTTATAGATGATGAANAATAATTTTACATTAACCGACAGCGTTTC	780
Qy	694	GTTATCCGTGGCCCAATGCGTGACCTCGCTCTGACCTGGTCGTAATATATGTTGATATCC	753
Db	781	GTTATTTGGAGACCCTCAAGAGAGATCGTGGTTTACTGACGCTAAATTAATGATTCATACG	840
Qy	754	TACGGCGCATGGCGGTCACGGTGGCGGTGCAATCTCTGCTGAAGATATCATCAAAATCG	813
Db	841	TATGGTGGTATATGCGGTCAATGGTGAGGCTGTTTTAGTGGTAAAGATCTACTAAGTA	900
Qy	814	GACCGTTCCGAGCTTACGACGACGCTTATGTCGCGAAACATCGTTGCTGCGCCTG	873
Db	901	GATGTTTACACACTTATGACGACMAATATGTAAGCTAAATAATTTGTCGACCTCGTTTA	960
Qy	874	GCGCATGTTGTGAATTCAGTTTCTTACGCAATGGCGTGGCGCAACCGCACTGCATC	933
Db	961	GCTAAACAATGTAGATACACTTGTGATATGCAATTTGGTGTACGACAAACCGTTTCCATT	1020
Qy	934	ATGCTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAAACACTGACCCTGCTGGTACT	993
Db	1021	TCAATTAATACGTTTGATGATCGGAAGAGTTTCAAGACGACGTTTATGTAACCTTAA	1080
Qy	994	GAGTCTTCGACCTCGCCCATACGGTCTGATTCAGATCTGATCTGCTGACCCGATC	1053
Db	1081	AAGGATTTTGAATTAAGACACGACGATATCATTAATAATGTTATGACTTAACAAACCGTA	1140
Qy	1054	TACAAAGAAGCCGACGATACGGTCACTTTGGTCTG-----GAACATTTCCCGTGGAA	1107
Db	1141	TATAGCAAAACAGACGCTATGGTCAATTTTGGACGTACAGACGCTATTTGTTACCATGGAA	1200
Qy	1108	AAAACGCAAGGCGGAGCTGCTGCCGATGCTG 1141	
Db	1201	AAATTAGATAAGTCAATGTTTTTAAAGATGCTG 1234	

```

RESULT 13
US-09-318-448-22
; Sequence 22, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:

```


APPLICANT: Johnson, William G.
 APPLICANT: Stenroos, Edward S.
 TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
 TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
 FILE REFERENCE: 601-1-057
 CURRENT APPLICATION NUMBER: US/09/318,448
 CURRENT FILING DATE: 1999-05-25
 NUMBER OF SEQ ID NOS: 46
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 22
 LENGTH: 1283
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-318-448-22

Query Match 28.8%; Score 332.2; DB 4; Length 1283;
 Best Local Similarity 57.3%; Pred. No. 5.9e-89;
 Matches 648; Conservative 0; Mismatches 468; Indels 15; Caps 2;

3 GCGCAAAACACCTTTTACGTCGAGTCGCTCTGAAGGCGATCTGACAAATTCGCTGA 62
 110 GGGCACATTCCTTTACCTCAGAGTCGGGGAAGGCCACCGCATTAAGATTGTGA 169
 63 CCAATTTCTGATGCGCTTTTACAGCGATCCTGCAACAGATCCGAAGCGCGTTGC 122
 170 CCAATCAGATGATGCTGCTCTGATGCGCCACTTACAGAGATCCTGATGCCAAGTAGC 229
 123 TTGCGAATCTACGTAACAAACCGCATGCTTTAGTTGGCGGAATTCACCCAGCGC 182
 230 TTGTGAACAGTTGCTAAACCTGGAATGATCTTCTGCGGGGAATTCATCTCAGAGC 289
 183 CTGGAGAGCATGGAAGATGATCAACCGTACACCGTTCGCGAATTCGCTGATTC 242
 290 TCGTGTGATACCAAGAACTGCTGTGAGCTGTTAAACATTTGATATATGATTC 349
 243 CGACATGCGCTTACGCTACCTCTGCGGTTCTGAGCGCTATCGGCAACAGTCTCC 302
 350 TTCCAAAGCTTTGACTACAAAGACTGTACGCTGCTGAGCTTGAGCAACAGTACCC 409
 303 TGACATCAACGAGCGCTTACCGTGC---GATCCCTGGAACAGGCGCGGTGACCA 359
 410 AGATATTGCTCAAGGTCTTATCTTGCAGAAATGAAAGAGATGCTGCTGAGACCA 469
 360 GCGTGTGATTTGGCTACCACTAATGAACGACGCTGCTGACGACACCTATCAC 419
 470 GCGCTTAATTTTGGCTATCCACATGATGAACGAGATGATGCTTTAACATTTCT 529
 420 CTATGACACCGCTCTGTGACAGCGTCAAGTGAAGTCCGTAACGCACTGCGCTG 479
 530 CTGTCACACCACTAATGCCAAACCTGCGAGAACTACGCCGTAATGCGACTTTGCCCTTG 589
 480 GCTGCGCGCGGAGCGGAACCGAGTGAATTTTCACTATGACAGCGCAAAATCGTTG- 538
 590 GTTACGCGCTGATTTCAAACTCAAGTACTGTGCACTAATATGACAGATCGAGCTGCTGT 649
 539 -----GATGATCTGCTGCTGCTGCTTTCACATCAGACGCTGTAAGAGATCGACCA 587
 650 GCTTCCCATCAGACTCCACCAATTTTATATCTCTTACACATCATCAAGAGCTTTTCTCT 709
 588 GAAATCGCTGCAAGACCGGTAATGAAGAGATCATCAAGCCAAATTCGCCCGCTGAATG 647
 710 TGATGAATGAGGATGCCCTAAAGAGAGATCATCAAGAGTGTGCTGCGGAATA 769
 648 GCTGACTTTCGCCCAAAATTTCTCATCAACCGCGGCTGCTTATCGTGGCC 707
 770 CTTGATGAGAGATTAACATCAACCTACAGCCAAAGTGGAGATTTTATTTGTTGGGCC 829
 708 AATGATGATGCGGCTGCTGATGCTGAATAATATGCTTATCTACGCGCGCATGGC 767
 830 TCAGGAGATGCTGCTTGTGATGAGCGGAATATATTGTGACACTTATGCGGTTGGG 889
 768 GCGTACAGTGGCGGCTGATCTCTGTGTAAGATCATCAAAATGAGCGGTTCCGAGC 827

DB 890 TCGTATGAGAGAGTGCCTTTTCAGAAAGATTATACCAAGTGCACCGTTACGTC 949
 828 CTAACGAGCAGCTTATGTCGCGAAGAAACATGCTGTGCTGCGCGCATGTTGTA 887
 DB 950 TTATGCTGCTGCTGAGGAGGCAAAATCTTGTGTAAGAGGCTGTGCGGAGGCTTC 1009
 888 AATTCAGTCTTCTACGCAATGCGGCTGAGTACACGCACTTCATCATGATGTAACATT 947
 DB 1010 TGTGAGCTCTCTTAATGATTTGAGATTCTCATCATTTATGATTCATTTCCATTA 1069
 948 CGGTACTGAGAAAGTCCCTTCTGAGACTGACCCCTGCTGATGATGATTTCTGACT 1007
 DB 1070 TGTGATCTCTCAGAGAGAGAGAGAGAGATTAAGATTTGAAAGAAATTTGATCT 1129
 1008 GCGCCCATACGCTGATTCAGATGCTGATGCTGATGCTGACCCGATCTACAAAGAACCC 1067
 DB 1130 CCGCCCTGCGGCTATTTGTAGGATTCGATCTGGAAGAACCAATTTATCAGAGCTCC 1189
 1068 AGCATACGCTCATCTTGTGCTGTAACATTTCCCGTGGGAAAAACCGACAA 1118
 DB 1190 AGCCTATGCGCACTTGTGTAAGGACAGCTTCCCATGGAAGTCCCAAAA 1240

RESULT 14
 US-09-318-448-20
 Sequence 20, Application US/09118448
 Patent No. 6210950
 GENERAL INFORMATION:
 APPLICANT: Johnson, William G.
 APPLICANT: Stenroos, Edward S.
 TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
 TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
 FILE REFERENCE: 601-1-057
 CURRENT APPLICATION NUMBER: US/09/318,448
 CURRENT FILING DATE: 1999-05-25
 NUMBER OF SEQ ID NOS: 46
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 20
 LENGTH: 3228
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-318-448-20

Query Match 27.4%; Score 316; DB 4; Length 3228;
 Best Local Similarity 56.8%; Pred. No. 5.8e-84;
 Matches 630; Conservative 0; Mismatches 465; Indels 15; Caps 2;

12 CTTTTTACGTCGAGTCCGCTCTCTGAAAGGCACTCTGACAAATTTGCTGACCAATTTTC 71
 125 CATGTTCACTACGAGCTGCTGTGGAGAGGACACCGCGATTAAGATCTGTGACCAATTCAC 184
 DB 72 TGATGCGGCTTTTACAGCGCATCTCTGAGAGAGATTCGGAAGACGCGGTGCTGGCAAC 131
 185 TGATGCAAGTCTGATGAGCCCATCTCAAGCAAGCCCAATGCGCAAGTGGCTGTGAGAC 244
 132 CTACGTAACAAACCGGATGCTTTAGTTGCGGAGAAATCAACCAACGCGCTGGGTAGA 191
 DB 245 AGTGTGCAACACCGCATGCTGCTGCTGTGAGATGATCACTCAATGCGCATGAGTGA 304
 192 CATGGAAGATCAACCGCTACACCGCTTCCGCAAAATTTGCTATGCTATTCACATGGG 251
 DB 305 CTACACAGCGGCTGTGAGGAGACCATCAAGCATATGCTGATGATGATGATGATGATG 364
 252 CTTTGAACGTAACCTCTGCTGCTGCTGCTGAGCGCTATGCGGCAACAGTCTCTGACATCAA 311
 DB 365 CTTTGAATCAACACTTGAACAGCTGCTGCTGCTTGGAGCACCAATCCCAATATTTGC 424
 312 CCAAGGCGTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
 DB 425 CCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484
 369 GTTGTGCTGACCACTAATGAAGACGAGTGTGATGCTGACGACCTATCATCTATGACCA 428

```

Db 485 GTTCGGTATGCTACGACGACAGAGAGATGCATGCCCTCACCATCATCTTGTCTCA 544
Oy 429 CCGCTGTGTACAGGCTAGGCTGAAGTGGCGTAAGAAAGGCACTTGGCGTGGCCGCC 488
Db 545 CAACCTGACGCCCGGATGGAGACCTCAGCGCTCGGCTCTCCCTCCCTGGCCGCC 604
Oy 489 GACGCGAAAGCCAGGCTGACTTTTCACT-----ATGACGACGCGAAATCGT 536
Db 605 TGACTCTAAGACTAGGTTGACAGTTCAGTACATGCAGACATGGCCAGTCACTCCCT 664
Oy 537 TGTATGATGCTGTCTGCTTTTCCACTCAGCAGCTCTGAAGATGACAGAAATCGT 596
Db 665 GCGCATCAGACCATCGTCACTCTGTGACAGACAGAAAGACATCAACGCTGGAGAGAT 724
Oy 597 GCAAGAACGGTAAATGCAAGAGATCATCAAGCCATTCTGCGCGCTGAATGGCTGCT 656
Db 725 GCGGAGGCGCTGAAAGAGACAGATCAGGCGCGTGGTGGCGCCCAAGTACCTGAGCA 784
Oy 657 TGCCACCAAAATCTTCATCAACCCGACCGGCTGTTGTTATGGTGGCCCAATGGGTA 716
Db 785 AGACACCGTCTACCACTGACGCGCAAGTGGCGTGTGCTCGAGAGTCCCGCAGGGCA 844
Oy 717 CTGCGTCTGCTAGCTGTGTAATAATTATCTGTATACCTAGCGCGCATGGCGCTCAGG 776
Db 845 TGGGGGTGCTACTGGCGGTAAGATTATTGTGACACCTATGGGGCTGGGGGCTCATGG 904
Oy 777 TGGCGGTGCTCTCTGTGTAAGATTCATCAAAAGTGGACCGTTCCCGACCTAGCGAC 836
Db 905 TGGTGGGCGCTTCTCTGGGAAGGCTACACCAAGGTAGAGCGCTCAGCTCATCTGCTG 964
Oy 837 ACCTTATGTGGCGAAACAATCTGCTGCTGGCGCTGGCGGATGCTGTGAATTCAGGT 896
Db 965 CCGCTGGGTGGCCAGTCTCTGTGTAAGAGAGGCTCTGCGGAGACTGCTTGTCCAGGT 1024
Oy 897 TTCTACGCAATCGCGCTGCTGTAACCGACCTCATCATGTAGAAATTTTGGTACTGA 956
Db 1025 TTCTATGCAATGCTGTGGCGGAGCGGCTGCTCATTTCCATCTTCACTACGGAACCTC 1084
Oy 957 GAAAGTCCCTTCTGACAACTGACCCCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 1016
Db 1085 TCACAAACAGAGAGAGAGTGTGATGTGTCATTAAGAACTTCCACCTCCGCGCGG 1144
Oy 1017 CGGCTGATTCAGATGCTGTGATGCTGACACCGCATCTACAAAGAACCGAGCATCGG 1076
Db 1145 CGTCAATGCTAGGATTTGCACTTGAAGAACCCATCTTACCAGAACAGCATGCTACGG 1204
Oy 1077 TCACCTTGTGCTGCAACATTTCCCGTGGCA 1106
Db 1205 CCATTTCGGAAGAGCATTTCCCATGGGA 1234

```

RESULT 15

```

US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: CDC 1551

```

```

: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: US-09-103-840A-2
: OTHER INFORMATION: represent a, t, c or g

```

```

Query Match 27.1%; Score 313.2; DB:4; Length 4403765;
Best Local Similarity 58.2%; Pred. No. 9,7e-82;
Matches 658; Conservative 0; Mismatches 418; Indels 54; Gaps 4;

```

```

Oy 13 CTTTATACGTCGAGTCCGCTCTGTAAG3GCATCTGACAAATTTGCTGACCAAAATTTCT 72
Db 156680 CTGTTTACAGTGAAGTCCGTGACAGAG3GACATCCCGACAAATCTGTGACCCATCAGC 1566739
Oy 73 GATGCCCTTTTAAACCGGATCTCTGACAGATGCCGAATCCGAACGCGCTTGTCTGGAACC 132
Db 156740 GACTCGTCTGTGACGCGCTTCTTAAAGCGGCAATTCACACACGCGCTGAGTGAAG 1566799
Oy 133 TACGTAAACACCGCATGTTTAAATTTAGTGGCGGCAATTCACACACGCGCTGAGTGAAG 192
Db 1566800 CTGGTACACACCGGAGAGTGCACGCTGTGGTGAAGTGAACCTCGGCTAAGAGCGG 1566859
Oy 193 ATCGAAGATCAC-----CCGTACACCGCTTCCGAATTTGCTATGTGCATTCC 243
Db 1566860 TTTGCCGACATGACACACAGGTCGCCGACGATCTCGAATCGGTACGACTGCTGG 1566919
Oy 244 GACATGGCTTTTACAGCTATCTCTGCGGCTTCTGACGCTATCCGCAACATCTCT 303
Db 1566920 GACAAAGTTTGTGACGCGGCGGACGCTGCG3GTAACATCGGATCGGCAAGTACCC 1566979
Oy 304 GACATCAACGAGCGGCTTGA-----CCGTCCCGATCCG 336
Db 1566980 GACATCGCCCAAGGCGCTGCACACCGCCACAGAGCCCGGCTGAGGCGCCGCTCATCCG 1567039
Oy 337 CTGGA---ACAGGCGCGGCTGACACAGGCTCTGATTTTGGTACGCAATTAATCAACC 393
Db 1567040 CTGGAATCTCCAGGCGCGCGGTGACACAGGCTCTGATTTTGGTACGCAATTAATCAACC 1567099
Oy 394 GAGTGTGATGGACAGCACTATCACTATGACACCGCTTGTGTACAGCTCAAGUCUGTA 453
Db 1567100 CCGGAATCTATCTCACTGCTGCTGCGCTGCGGCTGCGGCTGCGGCTGACCGAG 1567159
Oy 454 GTGCGTAAACAGGCACTCTGCGTGGCTGCGCGGACCGGACGGAAGACAGTCAATTTT 513
Db 1567160 GTGCGAABAAGGCGGTGCTGCGCTACCTGCTGCGGATGGAAGACGAGTCACTATTC 1567219
Oy 514 CAGTATACGACGCAAAATGTTGTATGATGCTCTGCTGCTTTCACATCAACACTCT 573
Db 1567220 GCTTACGAGAGCAACTCTCGGTGCGGCTG3GATACGTTGATCTCCACCAGCACCGG 1567279
Oy 574 GAAGAGATGACCAAGAAATGCTGCAGAAAG3CGGTAAATGGAAGATCAATCAAGCCAAAT 633
Db 1567280 GCCGATATCGACCTGAGAAAGCGTTGATCCCGACATCCGGGAAAGGTGCTCAACACC 1567339
Oy 634 CTGCCCCGCTGA-----ATGCTGACTTCTGCCACCAAAATTTCTTATCAAC 678
Db 1567340 GTGCTGTGACAGCACTGCGCCACCAAAACCTTGACGCTGACGCTGCGGGTGTGTGAAC 1567399
Oy 679 CCGACGGGCTGTTCTGTTATGCTGCGCCAAATGGGTACGCTGCTGACTGTCTGTA 738
Db 1567400 CCGACCGGCAAGTCTGCTGCGCGGCGCGATGGCGAATGGCGGCTCACCGGCCCAAG 1567459
Oy 739 ATTATGTTGATCTACAGCGGCGATGGCGGCGGATGGCGGCTGATCTGTTGTTAA 798
Db 1567460 ATCATCTGTGACACCACTGACGCGGCTGGCGCCCGACGCGGCGGCGGCTTCTCCGCAAG 1567519
Oy 799 GATTCATCAAAAGCTGACCGTTCGCGAGCTTACGACAGCATGTTATGTCGGAAGAAATC 858
Db 1567520 GATTCGCTCAAGATGAGACGGTGGCGGCTTACGATGCTGCTGGTGGCAAGATCTC 1567579
Oy 859 GTTGTCTGTGCTGCGGCGGCTGTTGTAATTCAGGTTTCTTACCAATCGCGTGTCT 918
Db 1567580 GTGCGCGCGGCTGCTGCTGTAAGCGGTGAGGTGCAAGTGTGCTTACGCTCGTAAAGCG 1567639
Oy 919 GAACGACCTTCATCATGTAGAAACTTTCGCTAGTGAAGAAAGTGTCTGTGACAAATTC 978

```

Db 1567640 GCACCCGTCGCCCTGTCGTCAGACGTTCCGTACCGAGACGAGACCCCGTCAAGATC 1567699
OY 979 ACCCTGCTGTAGTGAGTTCTTCGACCTCGCCCATACGGTGTGATTGATGCTGGAT 1038
Db 1567700 GAGAGGCCATCGCGAGGTATTGACCTCGCCCGGTGCCATCATCCGACCTGAAC 1567759
OY 1039 CTGCTGACCCGATCTACAAGAAACCGACATACGTCACCTTGTGTCG 1088
Db 1567760 CTGTTGCCGCCGATCTATGCGCCGACCGCCCTACGGGCACTTCGGCCG 1567809

Search completed: April 22, 2003, 14:26:12
Job time : 4798.39 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 12:23:08 ; Search time 147.738 Seconds

(Without alignments)
8506.861 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155

Sequence: 1 atggcaaaacacccctttac.....atgcctccgcgtgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US05_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC105_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1155	100.0	1155	10	US-09-815-242-6198 Sequence 6198, Ap
2	926.8	80.2	1167	10	US-09-815-242-6641 Sequence 9641, Ap
3	641.4	55.5	1155	10	US-09-815-242-7099 Sequence 7099, Ap
4	586.2	50.8	1191	10	US-09-815-242-7694 Sequence 7694, Ap
5	391.4	33.9	1203	10	US-09-974-300-972 Sequence 972, App
6	375.4	32.5	1191	10	US-09-815-242-4516 Sequence 4516, Ap
7	375.4	32.5	1296	10	US-09-815-242-8183 Sequence 8183, Ap
8	375.4	32.5	1296	10	US-09-815-242-8758 Sequence 8758, Ap
9	375.4	32.5	10320	7	US-08-781-986A-143 Sequence 143, App
10	374	32.4	4848	9	US-09-943-702-1 Sequence 1, App1
11	374	32.4	4848	9	US-09-943-702-4 Sequence 4, App1
12	374	32.4	4848	9	US-09-943-702-6 Sequence 6, App1
13	370.6	32.1	1185	10	US-09-815-242-6669 Sequence 6669, Ap
14	369.8	32.0	1191	10	US-09-815-242-9315 Sequence 9315, Ap
15	369.2	32.0	1158	10	US-09-815-242-3823 Sequence 3823, Ap
16	368.2	31.9	1191	10	US-09-815-242-9493 Sequence 9493, Ap
17	368.2	31.9	1191	10	US-09-805-847-1 Sequence 1, App1
18	356.6	30.9	1377	10	US-09-070-927A-672 Sequence 672, App
19	331.8	28.7	1221	9	US-09-738-626-1773 Sequence 1773, Ap

20	331.8	28.7	1344	9	US-09-746-660A-119 Sequence 119, App
21	314.4	27.2	1487	10	US-09-880-107-1655 Sequence 1655, Ap
22	309.8	26.8	1185	10	US-09-887-576-784 Sequence 784, App
23	305.6	26.5	1190	10	US-09-887-576-807 Sequence 807, App
24	296.8	25.7	1198	10	US-09-815-242-7386 Sequence 7386, App
25	292	25.3	1190	10	US-09-887-576-776 Sequence 776, App
26	285	24.7	1182	10	US-09-887-576-4501 Sequence 4501, App
27	264.8	22.9	1173	9	US-09-938-842A-2445 Sequence 2445, Ap
28	210.8	18.3	609	10	US-09-974-300-5505 Sequence 5505, Ap
29	208	18.0	2774	10	US-09-802-853-5 Sequence 5, App1
30	162	14.0	495	10	US-09-887-576-800 Sequence 800, App
31	155.2	13.4	495	10	US-09-887-576-787 Sequence 787, App
32	129	11.2	442	10	US-09-770-444-871 Sequence 871, App
33	120	10.4	366	9	US-10-076-622-278 Sequence 278, App
34	120	10.4	366	10	US-09-604-287A-278 Sequence 278, App
35	120	10.4	366	10	US-09-339-338-278 Sequence 278, App
36	120	10.4	366	12	US-10-007-805-278 Sequence 278, App
37	119.2	10.3	379	10	US-09-878-574-3276 Sequence 3276, Ap
38	114.6	9.9	411	10	US-09-878-574-3928 Sequence 3928, Ap
39	113.4	9.8	383	10	US-09-878-574-1980 Sequence 1980, Ap
40	110	9.5	413	10	US-09-878-574-3703 Sequence 3703, Ap
41	105.6	9.1	393	10	US-09-878-574-3731 Sequence 3731, Ap
42	105.2	9.1	394	10	US-09-878-574-1654 Sequence 1654, Ap
43	104	9.0	389	10	US-09-878-574-2715 Sequence 2715, Ap
44	102.2	8.8	386	10	US-09-878-574-1498 Sequence 1498, Ap
45	102.2	8.8	389	10	US-09-878-574-1447 Sequence 1447, Ap

ALIGNMENTS

RESULT 1
US-09-815-242-6198
Sequence 6198, Application US/09815242
Patent No. US200206156A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 6198
LENGTH: 1155
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1155)
US-09-815-242-6198

Query Match 100.0%; Score 1155; DB 10; Length 1155;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 ATGGCAAAACACCTTTTACGTCGAGTCCGTCTCTGAAAGGACATCCGACAAATTTGCT 60
    |||||||
Db 1 ATGGCAAAACACCTTTTACGTCGAGTCCGTCTCTGAAAGGACATCCGACAAATTTGCT 60

Oy 61 GACCAAAATTTCTGATGCCGTTTATAGAGCGATCCTCGAAGAGATCCGAAAGCAGCGCTT 120
    |||||||
Db 61 GACCAAAATTTCTGATGCCGTTTATAGAGCGATCCTCGAAGAGATCCGAAAGCAGCGCTT 120

Oy 121 GCTTGGCAAAACCTAGTAAACCGCGATGTTTTAGTTGGCGGGAATATACCACCGC 180
    |||||||
Db 121 GCTTGGCAAAACCTAGTAAACCGCGATGTTTTAGTTGGCGGGAATATACCACCGC 180

Oy 181 GCCTGGTAGATCGAAGAGATCACCGGTACACCGTTCCGCAAAATTTGCTATGTGAT 240
    |||||||
Db 181 GCCTGGTAGATCGAAGAGATCACCGGTACACCGTTCCGCAAAATTTGCTATGTGAT 240

Oy 241 TCCGACATGGGCTTTGACGCTACTCTGCGGTTCTGACGCTATCGCAACAGTCT 300
    |||||||
Db 241 TCCGACATGGGCTTTGACGCTACTCTGCGGTTCTGACGCTATCGCAACAGTCT 300

Oy 301 CCTGACATCAACGAGGCGTTGACGCGATCCGCGTGAACAGGCGGCGGTGACGAG 360
    |||||||
Db 301 CCTGACATCAACGAGGCGTTGACGCGATCCGCGTGAACAGGCGGCGGTGACGAG 360

Oy 361 GGTGTGATGTTTGGCTACGCACTAATGAACCGACGCTGATGCGACGACTATGACC 420
    |||||||
Db 361 GGTGTGATGTTTGGCTACGCACTAATGAACCGACGCTGATGCGACGACTATGACC 420

Oy 421 TATGACACGCTCTGATACAGCGTGAAGCTGAGTAAAGGCGACTCTGCGGTG 480
    |||||||
Db 421 TATGACACGCTCTGATACAGCGTGAAGCTGAGTAAAGGCGACTCTGCGGTG 480

Oy 481 CTGGCGCGGAGGGAAGGAGGATGACTTTGAGTATGACGAGCGCAAAATGTTGGT 540
    |||||||
Db 481 CTGGCGCGGAGGGAAGGAGGATGACTTTGAGTATGACGAGCGCAAAATGTTGGT 540

Oy 541 ATCGATGCTGTGCTGCTTTCCACTCAGCACTCTGAGAGATCGACGAGAAATCCGTCGA 600
    |||||||
Db 541 ATCGATGCTGTGCTGCTTTCCACTCAGCACTCTGAGAGATCGACGAGAAATCCGTCGA 600

Oy 601 GAACCGGTAATGAAGAGATCATCAAGCCAAATTTGCCCGCTGAATGGCTGACTTGGCC 660
    |||||||
Db 601 GAACCGGTAATGAAGAGATCATCAAGCCAAATTTGCCCGCTGAATGGCTGACTTGGCC 660

Oy 661 ACCAAATTTCTCATCAACCGGAGCGGTCTGTTATCGGTGGCCCAATGGGTGACTGC 720
    |||||||
Db 661 ACCAAATTTCTCATCAACCGGAGCGGTCTGTTATCGGTGGCCCAATGGGTGACTGC 720

Oy 721 GGTGTGACTGCTGCTAAATATGTTGATACCTACGCGGCGATGCGCGTACAGGTGAC 780
    |||||||
Db 721 GGTGTGACTGCTGCTAAATATGTTGATACCTACGCGGCGATGCGCGTACAGGTGAC 780

Oy 781 GGTGCAATTTCTGTGTAAGATCATCAAAAGTGAACCGCTTCCGACGCTACGACAGCT 840
    |||||||
Db 781 GGTGCAATTTCTGTGTAAGATCATCAAAAGTGAACCGCTTCCGACGCTACGACAGCT 840

Oy 841 TATGTGCGAAACATCGTGTGCTGCGCTGGCCGATCGTTTGAATTCAGATTTCC 900
    |||||||
Db 841 TATGTGCGAAACATCGTGTGCTGCGCTGGCCGATCGTTTGAATTCAGATTTCC 900

Oy 901 TACCAATTCGCGTGTGCTGAACGACGCTCATCATGTGTAAGATTTGCTGACTGAGAAA 960
    |||||||
Db 901 TACCAATTCGCGTGTGCTGAACGACGCTCATCATGTGTAAGATTTGCTGACTGAGAAA 960

Oy 961 GTGCGTTCTGTGAACATGACCTGCTGTGATGAGTTCTTGAACCTGCGCCCATACGGT 1020
    |||||||
Db 961 GTGCGTTCTGTGAACATGACCTGCTGTGATGAGTTCTTGAACCTGCGCCCATACGGT 1020
  
```

```

Oy 1021 CTGATTCAGATGCTGATCTGCTGACCGGATCTACAAAGAACCGACGATAGGCTCAC 1080
    |||||||
Db 1021 CTGATTCAGATGCTGATCTGCTGACCGGATCTACAAAGAACCGACGATAGGCTCAC 1080

Oy 1081 TTTGCTGCTGATCAATTTCCGTTGGGAAAAAACCGACAAGCGCAGCTGCTGGCGATGCT 1140
    |||||||
Db 1081 TTTGCTGCTGATCAATTTCCGTTGGGAAAAAACCGACAAGCGCAGCTGCTGGCGATGCT 1140

Oy 1141 GCCGCTGTGAAGTAA 1155
    |||||||
Db 1141 GCCGCTGTGAAGTAA 1155

RESULT 2
US-09-815-242-9641
; Sequence 9641, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9641
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1167)
; NAME/KEY: misc_feature
; LOCATION: (1)..(1167)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9641

Query Match 80.2%; Score 926.8; DB 10; Length 1167;
Best Local Similarity 87.6%; Pred. No. 11e-296;
Matches 1012; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
  
```

```

Sequence 7099 Application US/09815242
Patient No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version: 4.0
SEQ ID NO 7099
LENGTH: 1155
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1155)
US-09-815-242-7099

Query Match          55.5%, Score 541.4, DB 10, Length 1155:
Best Local Similarity 72.2%, Pred. No. 4.5e-202:
Matches 834, Conservative 0, Mismatches 321, Indels 0, Gaps 0:

OY 1 ATGCGAANAACCTTTTACGTCGAGTCGCTCTGTGAAGGGCATCTCGACAATTTGCT 60
DB 1 ATGCTTACTTATTTATTTATTCATCTGAATCAGTCTCTGAAGGCAATCCAGATTAATTTCC 60
OY 61 GACCAAAATTTTGATGACCGCTTTTAGACGGATGCTCTGAAACAGACATTCGAAAGCAGCGCTT 120
DB 61 GATCAAAATTTTGTATGCGGTACTTATGAAATCTCTTAAACAAAGATTCCTCAAAAGCAGCAGTA 120
OY 121 GCTTGCGAACCTTACGTAAAAACCGGATGCTTTTAACTTGGTGGCGGCAATTCACCAACGAC 180
DB 121 GCTTGTGAACCTTACGTAAAAACCGTATGATGCGATTTGGATATTTGGTATTAATGAACAT 180
OY 181 GCGTGGGTGAGACATCGAAGAGATCAACCCGTACACCGTTCCGGAATTTGGCTATGTGCTAT 240
DB 181 GCGTGGGTGATATTTGAAGAAATTTAAACCGGTAAAGATTTGGATATTTGGTATTAATGAACAT 240
OY 241 TCCGACATGAGCGCTTGACGCTACCTCTGTGCGTTTGTAGCGATCGCATCGCAAAACAGTCT 300
DB 241 TCTGAAATGGCGTTTGATGATGCTATCTTGTGTGTAAGCTTAATTCGATTTGGTAAACATGCT 300
OY 301 CCTGACATCAACGAGGCGTTGACCGGTGCGATCCGCTTGAACAGAGGCGCGGCTGACGAC 360
DB 301 GCAGCATATTTAATCAAGCGGTTGATGCTGTAATTCATTAGATCAAGCGCCAGCGCAACAA 360
OY 361 GGTGATGATTTGGTGTATGCGCAACATGAACGAGATGATTAATGCTGTGACGACCATTTACT 420
DB 361 GGTATTTATGCTTTGGTGTATGCGCAACATGAACGAGATGATTAATGCTGTGACGACCATTTACT 420

```



```

Db 675 CATCAACCCGACCGGCACTTCGTGATCGCGCGCCGCTCGGCGACTCGGCTGACCG 734
Oy 732 TGGTAATATTCGTGATACCGACGGGCAATGGCGGCTCAACGGGCGGTCATTC 791
Db 735 ACCGAATGATCTGTCGATCTCAGCGGCGCATGGCCCGCCACGGCGGCGGCTTC 794
Oy 792 TGGTAAGATCCATCAAAAGTGAACCGCTCCGACGCTACGAGCACTTATTCGCGAA 851
Db 795 CGGCAAGACCGGCTCAAGAGTGCAGCGCTCCGCGCTACGCGCGGCTAGTGCGAA 854
Oy 852 AAGCATCTGCTGCTGCGGCTGCGCGCATGCTTGTGAATTCAGGTTTCTTACGCAATCG 911
Db 855 GAACTACCTGCGCGCGCGCTGCGAGGCTTGCAGAGTTCAGGTTTCTTACGCAATCG 914
Oy 912 CGTGGGTGAACCGGCTGCTGATGTTAGAACTTCGCTGCTGAGAAAGTGGCTTCTGA 971
Db 915 CGTGGCGGACCGGCTGCTGATGTTAGAACTTCGCTGCTGAGAAAGTGGCTTCTGA 974
Oy 972 ACAACTGACCTGCTGCTGATGTTAGAACTTCGCTGCTGAGAAAGTGGCTTCTGA 1031
Db 975 GAAATGCTGTCGAACGTGCTCGGAGAGCATTCGACCTGCGCCGATACGCGATCAACAGAT 1034
Oy 1032 GCTGGATCTGCTGACCGCATCTACAAAGAACCGGATACGCTTGTGCTGTA 1091
Db 1035 GCTGCGACCTGCTGACCGCATCTACAAAGAACCGGATACGCTTGTGCTGTA 1094
Oy 1092 ACATTTC 1098
Db 1095 TCCGTTT 1101

```

RESULT 5

```

US-09-974-300-972
; Sequence 972, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974.300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 972
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-972

```

Query Match 33.9%; Score 391.4; DB 10; Length 1203;

Best Local Similarity 60.9%; Pred. No. 3.6e-119; Matches 713; Conservative 0; Mismatches 416; Indels 42; Gaps 3;

```

Oy 14 TTTTACGTCGACGCTCTCTGAAAGGCAATCGTGAACAAATTCGTAACAAATTCG 73
Db 20 TATTTACATCAGATCAGTACGAGGAGCATCCGCAAAATTCGACCAAGATTCG 79
Oy 74 ATGCCCTTTTACAGCGATCTCGAAGCAGATCCGAAACAGCGCTTGTGCGAAGCT 133
Db 80 ATAGTATATTAAAGCAATTTTAAAGAAAGATCCGCAATCCGCTTGTGCTGAACAT 139
Oy 134 ACCTAAACCGGATGTTTGTGCGGCAATTCACACACGCGCTGGTGAACA 193
Db 140 CTTGTGAACCGGATGTTGCTGCTGTAAGCGGTGAATTCACGACGCTCAATACGTGATA 199
Oy 194 TCGAAGATCAACCCGTAACCGCTTTCGGAATTTGGCTATGTGATTCGACATGGCT 253

```

```

Db 200 TTCCGAAAACAGTCCCGGAAACGATTAAAGAAATTCGCTACACAGTGCAGAAATACGCTT 259
Oy 254 TTGAGCTTAACCTCGTGGGCTTGAGAGTATGGGCAACAGTCTCTGACATCAAC 313
Db 260 TTGATGCGGAAATTCGCGGCTTATACCTTATAGATGAACAGTCTCCAGATTTGCA 319
Oy 314 AGGCGTTGAC-----GTGCGATTCGCTG 340
Db 320 TGGAGTGCACGAGCGCTTGAAGCGCGTGAAGAGCAATGAGCGACGCTGAATTTGAAG 379
Oy 341 AACAGGCGCGGCTGACAGGCTCTGATGTTTGGCTACGCACTAATGAACCGACGTC 400
Db 380 CGATGCGAGCCGAGACCAAGGTTTAAATGCTGTTTCCCTGCAACGAAACAAAGAAC 439
Oy 401 TGATGCGAGCTTATCAACCTATGACACCGCTGCTGATGAGCTGACGCTGAGAGTGCCTA 460
Db 440 TTATGCGCTGCGATTCGCTTGGCCATTAATTTGCTGCGCTTGAAGAGAGTCCGA 499
Oy 461 AAAACGCGACTCTGCTGCTGCTGCGCGCGGACGCAAAAGCAGGTGACTTTTCACTATG 520
Db 500 AAGAGGAATTTGCGGTATCTGCGTCCGAGGAAAGACAGGTAACGTTGATGATG 559
Oy 521 ACGACGCAAAATCG---TTGCTATGATCTGCTGCTTTCACCTCAGACTTGAG 577
Db 560 ATGAGAACAAACAGCGATTCGATGATCAATCTCTCATTCAGCTCAGCAATCACTG 619
Oy 578 AGATGACCAAGAAATGCGTCAAGAAAGCGTATGGAAGATTCATCAAGCCCAATTCCTC 637
Db 620 AGATTTCTGTTGACAAATTCACGCGCAATCTGAAGAGACGCTCATCAATTCGATGTT 679
Oy 638 CCGCTGAATGCTGACTTCTGCCACCAATTTCTATCAACCGACGCGCTCTTCTTGA 697
Db 680 CAANAAGAGCTGATGATGAACACGAAATCTTATTAATCTTACGCGGCTTTGCTTA 739
Oy 698 TCGTGGCCCAATGGGTGACTGCGCTGCTGATGCTGTAATTAATTCGTTGATACCTTACG 757
Db 740 TCGGGGACCGGAGGGGATGCGGTTTAAACGGCGCAAAATTCATCTGTTGACCTATG 799
Oy 758 GCGGATGGCGGCTGACGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 817
Db 800 GCGGCTATGCGCGCTGCGGAGGAGGCTTTTCCGCAAGATCGACAAAGGTTGATA 859
Oy 818 GTTCGCAACCTACGACGCTTATGTCGCAAAATTCGCTGCTGCTGCTGCTGCTGCTGCTG 877
Db 860 GATCGACCTTATGACGCGAATATGTCGCAAAATTCGCTGCTGCTGCTGCTGCTGCTGCTG 919
Oy 878 ATGCTTGAATTCAGCTTCTTACGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 937
Db 920 ATTCCTGTGAATACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 979
Oy 938 TGAAGACTTGGCTGCAAGAGGCTTGAACCACTGACCTGCTGCTGCTGCTGCTGCTGCTG 997
Db 980 TCGATACATTCGCTGACGCAAGGAGTGAAGAACATTTGACAGTGTTCGCAAAA 1039
Oy 998 TCTTGACCTGCGGCAATACGCTGATTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1057
Db 1040 ACTTTGATTAAGACTGCGGCAATTAATAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1099
Oy 1058 AAGAAACCGGACATACGCTCACTTTGCTG-----TGACATTTCCCGTGGA AAAA 1111
Db 1100 AACAAACCGGACATACGCTCACTTTGCTG-----TGACATTTCCCGTGGA AAAA 1159
Oy 1112 CGGACAAAGCGAGCTGCTGCGGATGCTG 1142
Db 1160 CAGACAAAGCAGATCAGCTGAAGAAAGATG 1190

```

RESULT 6

```

US-09-815-242-4516
; Sequence 4516, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```

```

: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4516
: LENGTH: 1191
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
US-09-815-242-4516

Query Match      32.5%: Score 375.4: DB 10: Length 1191:
Best Local Similarity 60.2%: Pred. No. 7.2e-114:
Matches 697: Conservative 0: Mismatches 421: Indels 39: Gaps 3:

OY 7 AAACACCTTTTACGTCGCGAGTCCGCTCTGGAAGGCATCCTGCAAAATTTGCTGACCAA 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13 AAACGATATTATTACTTCAAGATCTGTTACAGACACCAGCAATAAATCGCTGACCAA 72
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 67 ATTCTGTATGCGCTTTAGAGAGCGATCCTCGAACAGATCCGAAGCAGCGGCTTGGC 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 73 GTGTACATGCAATATTAGATGCTATTATTAAGACAGCAACCAATGCAACGCTGCTGT 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 127 GAACACTGCTAAACACGCGCATGTTTACTGTGCGCGGAATTCACACCGCGCTGG 186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 GAACAACTGTTACACAGATGATGGCAATTCGCGCGGAATTTCTACAAACAATAT 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 187 GTAGACATCGAAGATCACCCGTTACACCGCTTGGCGAATTTGCTATGCTATCCGAC 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 GTTATATTCCGAAGATGTTAGAGAAACAATCAAGAAATTTGATGATCAACAAGACAAA 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 247 ATGGGCTTGAAGCTAATCCTGTGCGCTGTGAGCGTATGCGCAAAACATCTCTGAC 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 253 TATGTTATGATTATGAACAATGCAATTTAACTGCGATTGATGAACAATCACTGAC 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 307 ATCAACGAGGCGTTGA-----CCGTGCGGATCCG 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 313 ATTGCAACAAGCGTGATGAAGCAATTAGATGATGATTAAGATAGCGAAGAAGAAATT 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 337 CTGAACAAGCGCGCGGTGACAGGCTGTGATTTGGCTACGCACTAATGAACGCGAC 396
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 373 GAAGCAACAGGTGACGAGCAAGGTTATGTTGGTTAGCGCAACAAGTAAAGAAAGAA 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 397 GTGCGATGCGACGACCTATCACTATCAACCGCTGTGATGACGCGGACGCTGAAGT 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 433 ACGTATATGCGCTGACCAATTTATTTATCATATCAATGCAAAAGCTTTATCTGATGTG 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 457 CTTAAACAGGCGACTCTGCGCTGCGCGCGCGAGCAAGCAACGACGATGCTTTTCAG 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 493 CGTAAAGATGCTAGCTTAATTTATTTAGACACAGACGCTAAAGTTCAAGCTGACGTTTGA 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY 517 TATGACG---ACGGCAAAATCGTTGGTATGATGCTGTGCTTCCACTGACACTCT 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 553 TATGATGAAATGATTAACCCGATACGATATGATACGATTTCTTATCAACACAACTGCT 612
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 574 GAAGATGACACGAGAAATGCGTCAAGAGCGGTTATGAGATATCATCAACCAAT 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 613 GAGATGCTGACCTTGACAAATTTCAAGAGACATCAAAAGCTATGCTATTTATCCACT 672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 634 CTCGCCGCTGAATGGCTGACTTCTGCCACCAATTTCTTCAACCCGACCGGCTGTTTC 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 673 GTTCGCAAAACCTTGATTAATGAACAACATAATTTCTACATCAATCTACAGGAGCTTT 732
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 694 GTTATCGGTGGCCCAATGCTGACTGCGTCTGCTGCTGTAATTTATGTTGATACC 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 733 GTATATCGGTGACCTCAAGCTGATGCGATTAACAGAGACTAAATCAATTTGTTACG 792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 754 TAGCGCGCATGCGCGCTGACGCGGCGGTCATTTCTGTTAAAGATCATCAAAAGTG 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 793 TAGCGTGCTATGACGCTCATGGTGGCGATGCTTCACTGTTAAAGATCTTACAAAGTA 852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 814 GACGCTTCCGACGCTACGCGACGCTTATGTCGCGAATAATCTGCTGCTGCGCTG 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 853 GACGCTTACGCTGATGACGCTGACGCTTATGTCGCTAAATATTTGTCGACGAGCTTTA 912
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 874 GCGGATGTTGTGAATTCAGGTTTCTTACGCAATGCGCGCTGCTGACGACCTGCATC 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 913 GCAGATCAATGTGAAGTACATTTGCGATATGCTGCTTCCAGAACCAATGTCATTT 972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 934 ATGCTAGAACTTTGCTGATGAGAAAGTGCCTTTCGACACGACCTGCTGTAAGT 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 973 GCAATTTATACATTTTGGACACAGTAAGTTTCTGAGAGACAACTTTCGACAAAGTA 1032
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 994 GAGTTCTTGACCTGCGCCCATAGCGTCTGATTCAGATGCTGATCTGCTGACACCGCATC 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1033 AAACACTTGAACCTGACGCTGACGATATTATTAATATGCTTGTGAACCAACGATT 1092
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1054 TACAAACAAACCGACGATACGCTTGTGCTGCTGTAAT-----TTCCCGTGGAA 1107
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1093 TACAAACAAACCTGCTTGTGCTGATTTGACGTTGACGATACAGATGCTTATTTCCATGGAA 1152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1108 AAAACGACAAAGCGCA 1124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1153 AAATTTAGCAAAAGTGA 1169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-815-242-8183
: Sequence 8183, Application US/09815242.
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-26
: PRIOR APPLICATION NUMBER: 60/253,625

```

```

: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8183
: LENGTH: 1296
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1296)
US-09-815-242-8183

```

```

Query Match      32.5%  Score 375.4; DB 10; Length 1296;
Best Local Similarity 60.2%  Pred. No. 7.5e-114;
Matches 697; Conservative 0; Mismatches 421; Indels 39; Gaps 3;

```

```

Oy 7 AACACCTTTTACGTCGCGAGTCGTCCTCTGCAAGGAGCATCTGCAAAATTTGTCGACCA 66
Db 115 AACACCTTTTACGTCGAGTCGTCGTCGACAGACACCCAGATTAATTCCTGACCA 174
Oy 67 ATTTCTGATGCGCTTTTACGCGCATCTCGAAGAGATCGCAAGCAGCGCTTCTTC 126
Db 175 GTGTCAGATGCAATATTAGATGCTATTAAAGACGACCAATTTGACAGCTTGT 234
Oy 127 GAACCTACGTAAACCGCGCATGTTTATGTCGCGGAAATCACCAACCGCGCTGG 186
Db 235 GAACACCTGTTCAAGAGGTATGCGCATGTCGCGGAAATTTCTCAACACATAT 294
Oy 187 GTAGACATGCAAGACATCCCGTAAACCGCTTCCGAAATTTGGCTATGCTATGCTCCGC 246
Db 295 GTTGTATTCCCAAGCTTTTAAAGAACATCAAGAAATTTGATACCAAGACCAAA 354
Oy 247 ATGGCTTTGACGCTACCTCTGTCGCTTTGAGCGCTATCGGCAACAGTCTCTGAC 306
Db 355 TATGCTTATGATTGAACAATGCAATTTTAACTGCGATGATGAACAATCACCTGAC 414
Oy 307 ATCAACCAAGCGCTTGA-----CCGTCCGATCCG 336
Db 415 ATTGCACAAGCGCTGATTAAGCATTTAGATCTGATTAAGATTAAGCAAAAGAAAT 474
Oy 337 CTGGAACAGCGCGGTGACAGGTCGATGTTGGTTAGCAACATTAAGAAACGCGC 396
Db 475 GAAGAACAGGTGACAGGTGACCAAGTTTAATGTTGGTTAGCAACATTAAGAAAGAA 534
Oy 397 GTGCTGATGCCAGCATCTATGACACACCGCTCTGATGACAGCTGAGCTGAAGTG 456
Db 535 AGCTATATGCTCTAGCATTTATTATACATCAATTTGGCAAAAGCTTATCTGATGTG 594
Oy 457 CTTAAACAGGACCTGCGCTGCGCGCGCGGACCGCAAAAGCAGCTGACTTTTCAG 516
Db 595 CTTAAAGATGCTAATTTATTACACACGAGTAAAGTTCAAGTCTGACTGTTGAA 654
Oy 517 TATGAGC---AGGGCAAAATGTTGATGATGCGTCTGCTTTCATGACAGCATCT 573
Db 655 TATGATGAATGATTAACCACTAGCTATGATGATGCTATCAACACATGCTT 714
Oy 574 GAAGAGATGACACAGAAATCGCTGCAAGAGAGCGTAAATGGAAGATCATCAAGCAAT 633
Db 715 GAGGATGTGACACTTGAACAATTTCAAGAAAGATCAAGATCAATGCTATTTATCAACT 774
Oy 634 CTGCGCGCTGAATGCTGCTCTGCGCACCAATTTCTATCAACCGACGCGCTTTTC 693
Db 775 GTTCAGAAATCTGATTAATGAACAACATAATTTCTACATCAATCTCAGACAGCTTT 834
Oy 694 GTTATGCTGCGGCAATGCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
Db 835 GTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894
Oy 754 TACGCGCGCATGCGCGCTGACGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813

```

```

Db 895 TACGTCGCTATGCAAGATGCTGCGCATGCTTCAAGTGAAGATCTTACAAAGCTA 954
Oy 814 GACCTTCCGACGCTTACGACAGCTTATGTCGCGGAAACATGCTGCTGCTGCTGCT 873
Db 955 GACCTTCCGACGCTTACGACAGCTTATGTCGCGGAAACATGCTGCTGCTGCTGCT 1014
Oy 874 GCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 933
Db 1015 GCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
Oy 934 ATGATGAACCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993
Db 1075 GCAATGATGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1134
Oy 994 GAGTCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1053
Db 1135 AACACCTTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194
Oy 1054 TACAAGAACCGGACGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1107
Db 1195 TACAACCAACGCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1254
Oy 1108 AACACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1124
Db 1255 AATTAAGACAAAGTGA 1271

```

RESULT 8

```

US-09-815-242-8758
: Sequence 8758, Application US/09815242
: Patent No. US20020061569A1

```

GENERAL INFORMATION:

```

: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsem, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8758
: LENGTH: 1296
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1296)
US-09-815-242-8758

```

```

Query Match      32.5%  Score 375.4; DB 10; Length 1296;
Best Local Similarity 60.2%  Pred. No. 7.5e-114;

```



```

Db 2504 TATGCTATGATTTGAAGCAATTTAACTCGATGATGAACAATCACTGAC 2445
Oy 307 ATCAACAGGCGCTTGA-----CCGTGCGGATCCG 336
Db 2444 ATTGCACAGAGCGGTGAAGCATTAAGATGATGATAAGATAGCAAGAAATTT 2385
Oy 337 CTGGAACAGGCGGCGGATGATGCTGATGTTGGCTACGCACTAATGAACGAC 396
Db 2384 GAACCAACAGGTGCGAGGTGACCAAGTTTATGTTGGTTACGCAAGCAATGAACAGAG 2325
Oy 397 GTGCTGATGCCAGCACTATCACCTATGACACCGCTGCTGTAACGCTCAGCTGAATC 456
Db 2324 ACGTATATGCTCTAGCAATTTATTTATCATCATCAATGCGCAAAACGTTTACGATCTG 2265
Oy 457 CGTAAACAGGCACTCTGCGCTGCTGCGCGCGGCGGAAACGCAAGTACTTTTCAG 516
Db 2264 CGTAAACAGTGTACCTTAATTTATTTACGACACAGCGTAAAGTTCAGCTCAGCTTGA 2205
Oy 517 TATGACG---ACGGCAAAATGCTGTGATGATGCTGCTGCTGCTTTCAGTACAGACTCT 573
Db 2204 TATGATGAATTAATTAACCCAGTACGTTATGATGATGCTGCTATCACACAGCATGCT 2145
Oy 574 GAAGATGACAGCAAGAAATCGCTGCAAGAACGCTAATGAAGATCATCAAGCAATTT 633
Db 2144 GAGCATGTGACACTTGAACAATTCAGAGACATTAACGACATGCTATTTATTCACACT 2085
Oy 634 CTGCGCGCTGAATGCTGACTTGTGCCCAAAATTTCTTATCAACCGGAGCTGCTTTC 693
Db 2084 GTTCCAGAAATCTGATTAATGAACAAATTAATTTCTACATCAATCCACAGGAGCGCTTT 2025
Oy 694 GTTATCGGTGCGCAATGGGAGTACGCTGCGCTGCTGCTAAATTAATGTTGATACG 753
Db 2024 GTATTCGCTGAGCTCAAGAGTATGACAGATTAACAGGACCTAAATATGTTGATGACG 1965
Oy 754 TACGCGCGCATGCGCGCTGACGCTGCGCGCTATCTCTGTAAAGATCATCAAAATG 813
Db 1964 TACGCTGCTATGACGCTGATGATGCGGAGTGTGCTGCTGTAAGATCTCAAAATG 1905
Oy 814 GACGCTTCCGACGCTACGACGACGATATGCTCGGAAACATGCTGCTGCTGCGCTG 873
Db 1904 GACGCTTCCGACGCTACGACGACGATATGCTGCTAAATTAATGTTGAGACGAGTTTA 1845
Oy 874 GCGGATCTGTGAATTCAGGTTTCTTACCAATCGCGCTGCTGACCAACGACCTCATC 933
Db 1844 GCAGTATCATATGAGATGATGCAATTTGCAATGCGATGCTGCTGCAAGCACTGATTT 1785
Oy 934 ATGTAGAACTTTGCTGATGAGAAAGTGCCTTCTGAACAACGACCTGCTGCTGATCT 993
Db 1784 GCAATTTGATACATTTGGAACAGGTAAAGTTCTGAAAGGCAAACTGTGCAAGCATGTA 1725
Oy 994 GAGTCTTTCGACCTGCGGCGCATAGGCTGATGATGCTGATGCTGCTGACCCGATC 1053
Db 1724 AAACCTTTGACCTTAAGACCTGACGATTTATTAATAAGCTGATTTGAACAAACGATT 1665
Oy 1054 TACAAGAAAGCGGACGATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1107
Db 1664 TACAAGAAAGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1605
Oy 1108 AAAACCGAACAGCGCA 1124
Db 1604 AAATTAAGCAAAAGTGA 1588

```

RESULT 10

US-09-943-702-1
 : Sequence 1, Application US/09943702
 : Publication No. US2003002322A1

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
 : Rostek Jr., Paul R.
 : TITLE OF INVENTION: SAM Operon
 : NUMBER OF SEQUENCES: 6

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/943,702
FILING DATE: 31-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/955,957
FILING DATE: 1997-10-22
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,072
REFERENCE/DOCKET NUMBER: P-10162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4848 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 986..2209
FEATURE:
NAME/KEY: CDS
LOCATION: 2241..3341
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-943-702-1
Query Match 32.4%; Score 374; DB 9; Length 4848;
Best Local Similarity 59.7%; Pred. No. 4,2e-113;
Matches 728; Conservative 0; Mismatches 420; Indels 72; Gaps 3;
Oy 1 ATGCGAAACACCTTTTACCTCCGAGTCTGCTCTGAAGGCAATCTGACAAATTTGCT 60
Db 986 ATGTACAGCTCCCGCTGTCACCTCGAGTCTGACCGGCGCACCCCGACAGATCGCC 1045
Oy 61 GACCAATTTTGTGATCCGCTTTTGAACGCGATCCCTGCAACGATGCCGAAGACAGCGCTT 120
Db 1046 GACCGGATCAGCGACCGCTCTGATCCCTGCTGCGCGGACCGCGGCGGCGGCGGCTC 1105
Oy 121 GCTTTCGCAACCTAGTAAACCGGCAATGCTTTTACTTGGCGGCAATGACCAACGAGC 180
Db 1106 GCGGTGAGACCTGATACCAACGCGTACGATGCTATGCGCGGAGAGTACCAACGACC 1165
Oy 181 GCTTGGTAGACATCGAAGACATCACCCGTAAACCGCTTCCGGAATTTGCTATGTGCAT 240
Db 1166 GCGTACGCGGCCATCGCCCAACTGTGTGCGCACACCGCTCTGCTCATCGCTACGACTGC 1225
Oy 241 TCCGACATGGGCTTTGACGCTAAGCTGCTGCGGCTTGTGAGCGGCTATGCGCAACAGTCT 300
Db 1226 TCGGCCAAGGCTTTCGACGGGCGCTGTCGCGGCTGTGCTGCTCATGCGCGGAGTCTC 1285
Oy 301 CCGTGCATCAACCAAGGCGCTTTCGAC----- 324
Db 1286 CCGGACATCGCGCGGCGGTGTGACACCGCGTACGACGCGGCGGCGGCGGCGGCGGCGG 1345
Oy 325 -----CGTCCGATCCCGCTGCAACAGCGCGCGGCTGACCGAGGCTGATGTTGGC 375

```


Db 1646 CCGCAUUCUGGAGCCGCCGCGGACUGGCGGACAGGCAUCCAGGACACC 1705
Oy 674 -----TCAACCCGACCGGTCTTCTTATCGTGGCCCAATGGT 714
Db 1706 GCGGGUACGGGUGUGUACUACCCAGCGGCGGUGUACAGUACGGCGGCGUAGG 1765
Oy 715 GACTGCGGTGACTGCTGCTAAATATATCTGATACCTAAGCGGCGCATGGCGCTAC 774
Db 1766 GACCGCGGUGUACCGCGGCGGCAUACUACGACACUACGCGGCGGCGCAU 1825
Oy 775 GGTGGCGGTGATCTCTGTGAAGATCAATCAAAAGTGACGCTTCCGAGCTAGCA 834
Db 1826 GCGCGCGGUGUACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1885
Oy 835 GCACGTTATGTCGCGAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894
Db 1886 AUGGCGUGGUGGCGGCAAGACGUGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1945
Oy 895 GTTTCCTACGCAATGCGGCGGCTGGAACCGACCTCCATCATGTAGAGAACTTTCGCT 954
Db 1946 GUGCGGUGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2005
Oy 955 GAGAAAGTGGCTTGTGAACACGACCTGCTGTAGCTGCTTCTGACCTGCGGCGCA 1014
Db 2006 GCGCAGGUGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2065
Oy 1015 TACGCTGTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
Db 2066 GCGCGGAGUACGCGGAGUACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2125
Oy 1075 GGTGACTTTGCTGCTGAC-----ATTTCGCGTGGGAAAAACCGACAAAGCGCTG 1128
Db 2126 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2185
Oy 1129 CTGGCGGCTGCTGCGGCT 1148
Db 2186 CUCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2205

RESULT 13
US-09-815-242-6669
Sequence 6669, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
PRIORITY FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6669
LENGTH: 1185
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)....(1185)
US-09-815-242-6669

Query Match 32.1%; Score 370.6; DB 10; Length 1185;
Best Local Similarity 60.0%; Pred. No. 2.8e-112;
Matches 688; Conservative 0; Mismatches 429; Indels 30; Gaps 3;

Oy 4 GCAAAACCTTTTACGTCGAGTCCGCTGCTGGAAGGCGATCTGCAAAATTCGTGAC 63
Db 7 GAAAGACATTTATTTACATGAGATGCGCTTTCAGAGGACATCCAGATAGTAGTGAC 66
Oy 64 CAATTTCTGATGCGGTTTATGACGCGATCTGGAACGAGATCCGAAGACGCGTTCCT 123
Db 67 CAAGTCAGTATGCAATTTTATGATGCGATTTTGGAGAAAGATCCGATGCGCGCGGCA 126
Oy 124 TCGGAACCTAGCTAAAAACCGGCGATGTTTATGTTGGCGGCAAAATCACACAGCGC 183
Db 127 TGTGAACATCCGTACACACTGGTTATGCTGCTATTTGGAGAAATTTCTACACAGCT 186
Oy 184 TGGGTACATCGAAGATGATCCCGTACACCGTTCGCAATTCGCTATGCTATTC 243
Db 187 TATGTAGATATCAAAATTTGTTGCAACATGTCGAAGATTTGCTATACAGCA 246
Oy 244 GACATGGCTTTGACGCTATCTGCTGCGGCTTCTGAGGCTATTCGCAACAGCTCT 303
Db 247 AAATATGATTTATGAGAGAGACTGTCGAGTCTTGTATCAATTCATGAGCAATCACT 306
Oy 304 GACATCAACGAGCGCTTTCGACCGTGC-----GATCCGCTGGA 342
Db 307 GATATCGCTCAAGGCTTACAGGAGCTTATGAGTGTGCTGATCAAGTGAAGAAAGCAT 366
Oy 343 CAGGCGCGGCTGACAGGCTGATGTTTGGTACGCAATTAAGCAACCGAGCTGCTG 402
Db 367 ATTGGCGCTGCTGACCAAGCTTACTGTTTGGTTTCCGTTGACGAACACGCAATTA 426
Oy 403 ATGCGACCTATACCTATATGACACCGCTGCTGACAGGCTGAGTGAAGTGTGCTAAA 482
Db 427 ATGCGATTCAGATCGGCTTATGATGATGATGATGATGATGATGATGATGATGATG 486
Oy 463 AACGCACTGCTGCGTGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 522
Db 487 GAAAGATCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 546
Oy 523 GAC--GCAAAATCGTTGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 579
Db 547 GATCAAGGTCAACACCAACCTGCTGATGATGATGATGATGATGATGATGATGATGAT 606
Oy 580 ATGACACGAATGCGTCAAGGCTGATGATGATGATGATGATGATGATGATGATGATG 639
Db 607 ACAACCTTGAACAAATGCAAAAGACATTAAGAAACAACTAATCAATGAGGTGATCT 666
Oy 640 GCTGAATGCTGATCTTTCGCAACCAATTCCTGCAACCGGCGGCTTTCCTTATC 699
Db 667 CATGAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 726
Oy 700 GGTGGCCCAATGCGTGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 759
Db 727 GGTGGCCCTCAAGAGATGCGGCGGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTG 786
Oy 760 GCGATGGCGGCTGACGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 819
Db 787 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846
Oy 820 TCCGACCTAGCAGGAGCTTATGCTGCGGCAAAACCTGCTGCTGCTGCTGCTGCTGCT 879
Db 847 TCTGCTAGTTATGCTGCGGCTTATGCTGCGGCAAAACCTGCTGCTGCTGCTGCTGCTG 906

Oy 880 CATTGTGAATTCAGCTTTCTTACCAATCGCGCTGGCTGACCGACCTCATCATGTGA 939
 Db 907 AAGTAAAGTCCAAATTAAGCTTAATGCAATGGCGTGGCGCAACGACGTCTATTCTATT 966
 Oy 940 GAACTTTGGCTAGTGAAGAAAGTCCCTTGTGACAACTGACCCCTGCTAGTACGATTC 999
 Db 967 AATACGTTTGTAGTACGACTGCTTCCAGAAAGTAATTAATGCAAGAGTTCGTAAGAAC 1026
 Oy 1000 TTTCACCTGGCCCATACGCTGCTGATTCACATGCTGATCTGCTGACACCCGATCTACAAA 1059
 Db 1027 TTGTATTATTACCTCCAGCAGGATTAATTAAGATTAATTAATTAATTAATTAATTA 1086
 Oy 1060 GAAACCCAGCATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1113
 Db 1087 CAACCTCGGCTTATGCTCATTTTGTGCAACAGATATTATTCATTACCGTGGAGCAACAA 1146
 Oy 1114 GACAAAG 1120
 Db 1147 GATTAAG 1153

RESULT 14 US-09-815-242-9315

; Sequence 9315, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haseldeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Travick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 9315
 ; LENGTH: 1191
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1191)
 ; US-09-815-242-9315

Query Match 32.0%; Score 369.8; DB 10; Length 1191;
 Best Local Similarity 60.6%; Pred. No. 5,1e-112;
 Matches 677; Conservative 0; Mismatches 407; Indels 33; Gaps 3;

Oy 14 TTTTACCTCGGAGTCCGCTCTGTGAAGGATCCTGACAAATTCGTCAGCAATTTCTG 73
 Db 17 TATTACCTGTGATCTGTATGTAGGGGATCCGATAGATTGACAGCAAAATTTTCAG 76

Oy 74 ATGCGTTTATAGACGATCTCTGACAGAGATCCGAAGACCGGTTTGTGCAAACT 133
 Db 77 ATGCGATTTTGAATGATATTTTGAAGAGATCCAGAGGGGACGTTGCTGCTGAACAG 136
 Oy 134 ACGTAAAGACCGGATGCTTTTGTGCGCGCAATCACCACCGGCTGGGTAGACA 193
 Db 137 CTGTATATACGTGTTCTGCTACGATTTTGTGTAATTTTCAAAATGCTATGTGATA 196
 Oy 194 TCGAAGAGATCACCCTTAACCGGTTGCGAATTTGCTATGTCATTCGACATGGCT 253
 Db 197 TTAAACGCTGTGTTGCTGATACATTCGACAGATTTGTTATACCAATACAGAAATATGAT 256
 Oy 254 TTGACGCTAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
 Db 257 TTTTCTGCTGACAGCGTGGAGATACACCACTTTTGTGGAATCTCTGCTGCTGCTGCT 316
 Oy 314 AGGCGTTGACCGTGCC-----GATCGCTGGAACAG- 345
 Db 317 AAGGTATTACGAAAGCTTGAGGTTGCGGAATGCTGATCAAGATCCACTGGACTTTGA 376
 Oy 346 --GGCGGCTGACAGGCTGATGTTTGGCTACGCACTAATGAACGAGCTGCTGA 403
 Db 377 TTGAGCAGGCTGACCAAGGCTCATTTTGTGATTCGATGATGAAGACGAAAGCCTTA 436
 Oy 404 TGGCAGCAGCTATACCTATGCAACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
 Db 437 TGGCATTGCTCAATTTGCTACGCTATGATTAATTTGCTGCTGCTGCTGCTGCTGCT 496
 Oy 464 ACGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523
 Db 497 CTGAGCAAAATTAATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 556
 Oy 524 AC---GGCAAAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 580
 Db 557 AAATGACGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 616
 Oy 581 TCGACCAAGATCTGCAAGAGCGCTAATGGAAGATCAACCAATTCCTGCGCG 640
 Db 617 CCACTAATGAACAAATCATCAAGATGATGATGCAAGGTCATCAAGAAAGATTATTCAT 676
 Oy 641 CTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 700
 Db 677 CTCTTATCTTGTATGATGAAGCAAAATTCCTTATTCATTCGACAGGCTGTTGTAAVCG 736
 Oy 701 GTGGCCAAATGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
 Db 737 GTGGCTCTCAAGGGGACTGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 796
 Oy 761 GCATGCGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820
 Db 797 GCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856
 Oy 821 CCGGAGCTTACGACGACGCTTATGTCGCGAAGAAATCATGTTGCTGCTGCTGCTGCTGCT 880
 Db 857 CAGCTCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 916
 Oy 881 GTTGTGAATTCAGCTTTCTTACCAATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
 Db 917 AGCGAAGTGCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976
 Oy 941 AAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
 Db 977 ATACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1036
 Oy 1001 TCGACCTGCGCCATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1060
 Db 1037 TTGACCTTCCCTCGCGAGGATTAATCCAAATGCGGACCTCAAGCGCTCAATTTACGCTC 1096
 Oy 1061 AAACCGGAGCATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
 Db 1097 AAACATGCTTACGGCTACATGCGAGCTACAGATAT 1133

```

RESULT 15
US-09-815-242-3823
: Sequence 3823, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlson, Karl L.
: APPLICANT: Zykind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815.242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-03-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 3823
: LENGTH: 1158
: TYPE: DNA
: ORGANISM: Enterococcus faecalis
US-09-815-242-3823

Query Match      32.0%; Score 369.2; DB 10; Length 1158;
Best Local Similarity 60.0%; Pred. No. 8e-112;
Matches 686; Conservative 0; Mismatches 428; Indels 30; Gaps 3;

OY 7 AACACCTTTTACCTCCGAGCTCTCTGTAAGGCATCCGACAAAATTCCTACCAA 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 AGACATTTATTTACATCAGAAATCCGTTTCAGAAAGACATCCGATAGCTACGACCA 60
OY 67 ATTTCGATGCCGTTTAAAGCCGATCCTCGAACAGATCCGAAGCAGCGCTTGTC 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 CTCAGTGAATGCAATTTTAGATCGCATTTTGGAGAAAGATCCGATGCCCGCTGCATGT 120
OY 127 GAAACCTGATAAAACCGCGCATGCTTTAGTTGGCGGCGGAATACCAACGACGCCCTGG 186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 GAAACATCCCGTAACAACCTGTTTACTCTGATTTGGAGAAATTTCTACACACCTTAT 180
OY 187 GTAGACATCGAAGATACCCGCTAACAACCGTTCGCGAAATGGCTATGTCATTCGCGAC 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 GTAGATATTCAAAATAATTTGTTGTCAAAATGTAAGATATTGTTATACAGACAAANA 240
OY 247 ATGGCTTTGACGTAACCTCTGTGCGGTTCTGACGCGTATCGCAAAAGTCTCTGAC 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 TATGATTTTGAATGGAAGACTGTGCGAGTCTTGTAGCAATGATGAGCAATCAGCTGAT 300
OY 307 ATCAACCAAGCGCTTGACCGTGC-----GATCCGCTGGAACAG 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 ATCGCTCAAGCGGTTGACGCGACCTTAGAAGTCCGTGATCAAGATGAATAAAGACGATATT 360
OY 346 GCGCGGGGAGACGAGGCTGTGATGTTGGCTAGCAACATAAGAACGAGCTGTGATG 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 GCGCGGTGGACCAAGGCTTACTGTTGGTTTGGCTTGACGAAACACCAATTAATG 420
OY 406 CCAGCACTATCACTATGACACACCGTCTGCTACAGCGTCAAGCTGAAGTGGTAAAAAC 465

```

```

DB 421 CCATTACCGATCGCGTTTAAGTCATCTGTTGGTTGGCCCTTTACGAGAAATTCGCAAGCAA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 466 GGCACCTCTGCCGTGGCTGCGCCGACCGCAAGAACCGAGGTCTTTCACTATGAGC-- 523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 AAGCTCTTACCTTACTTACGTCTGATGCAAAATCTCAAGTAAACAGTGAATATGATGAT 540
OY 524 -ACGCGAAATCGTTGGTATGATGCTGCTGCTTCCACTACGACACTCGAAGAGATC 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 541 CAAGTCAACCAACAGAGTGTGATACGATTTGTCATCACTACGACAGATGATGAAGCA 600
OY 583 GACCAAGAAATCGCTGCAAGAAAGCGGTATGAGAGATCATCAACCAATTCGCGCGCT 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 601 ACCTTAGAACAAATCGAAAAAGACATTTAAAGAACAGTAATCAATGAGTATTCCTCAT 660
OY 643 GAATGGCTGACTTCTGCGACCAATCTTCATCAACCGGACGGTCTGTTGTTATGGT 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 GAACCTTTAGATGATGAACAAATACCTTTATCAACCACTGGTCCGCTTGTATCGGT 720
OY 703 GCGCCCAATGGGTGACTGGGTCTGACTGTGCTAAATTAATTCGTTGATACCTAGCGCGC 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 GCGCCTCAAGAGAGATGCGGCTTAACTGTGTGTAATAATTAATGTTGATACCTTATGCGCT 780
OY 763 ATGGCGCTCACGGTGGCGGTGATCTCTGTGTAAGATCCATCAAAAGTGAACGCTTCC 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 TATGCTGCTCATGGCGGGGTGCTTTTCTGAAAAAGATGCAACCAAGTTGACCGCTCT 840
OY 823 GCAGCTACGAGCAGCTTATGTGCGCAAAACATCGTTGCTGCTGCGCTGCGCGATCGT 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 841 GCTAGTTATGCTGCGCGCTTATATCGCGAAAAACATTTGCTGCTGCTGCTGCGCAAAAAA 900
OY 883 TGTGAATTCAGGTTTCTTACGCAATCGCGCTGCTGCAACCGACTCCATCATGTGTA 942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 901 GTAGAGTCCAAATTAAGCTTATGCAATTTGGCTGGCGCAACGATCTATTTCTATTAAT 960
OY 943 ACTTCCGCTAGCAAGAGTCCCTTCTGAACACTGACCCCTGTGCTAGCTAGTCTTC 1002
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 961 ACGTTTGTAGAGAGTCTCTTCCAGAAAGTAAATTAATGCAAGCAGTTCTGTAATAACCTT 1020
OY 1003 GACCTGCGCCCATACGCTCTGATTCAGATGCTGATGCTGCAACCGCATCTACAAAGAA 1062
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1021 GATTTACCTCCAGCGATATTATGAAATGCTAGATTACCTGCCCAATTTATTAACAA 1080
OY 1063 ACCGAGCATAGCGGTCACTTTGTCG-----TGAACTTCCGCTGGGAAAAAACCGAC 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1081 ACTGGGCTTATGGTCATTTTGTGTCGAACAGATATTGATTCGTTGGAACAAACAGAT 1140
OY 1117 AAAG 1120
    ||||
DB 1141 AAAG 1144

```

Search completed: April 22, 2003, 15:13:25
 Job time : 188.738 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:05:34 ; Search time 1882.93 seconds

(without alignments)
9934.410 Million cell updates/sec

Title: US-09-622-419-1

Perfect score: 1155

Sequence: 1 atgcacaaacaccttttctac.....atgcgcgcgctcgaagtaa 1155

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
EST :
1: em_estda :
2: em_esthum :
3: em_estlin :
4: em_estnu :
5: em_estrov :
6: em_estrpl :
7: em_estro :
8: em_htc :
9: gb_est1 :
10: gb_est2 :
11: gb_hic :
12: gb_estc3 :
13: gb_estc4 :
14: gb_est5 :
15: em_estfun :
16: em_estom :
17: gb_gss :
18: em_gss_hum :
19: em_gss_inv :
20: em_gss_pln :
21: em_gss_vit :
22: em_gss_fun :
23: em_gss_mam :
24: em_gss_mus :
25: em_gss_other :
26: em_gss_pro :
27: em_gss_rtd :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.4	23.5	1912	11	AY109333 Zea mays
2	248	21.5	994	14	BM807271 AGENCOURT
3	246.6	21.4	920	17	BH770810 LLMGtag955
4	235	20.3	889	17	AL405027 TT end of
5	233.2	20.2	1056	10	BE035235 M002F06 M
6	232.8	20.2	1060	10	BE412603 MCG002.B0

Result No.	Score	Query Match	Length	DB ID	Description
7	225.2	19.5	1045	14	B0071434
8	219.2	19.0	921	17	CNS070VU
9	219.2	19.0	1071	14	B0055329
10	214.4	18.6	977	14	B0964578
11	209.6	18.1	768	12	BF625331
12	209	18.1	755	13	BF1921837
13	207.6	18.0	817	12	BC414252
14	207.4	18.0	720	13	BJ432712
15	207.4	18.0	725	13	BJ432712
16	207.4	18.0	743	13	BJ428430
17	207.4	18.0	750	13	BJ428525
18	207	17.9	798	12	BF629512
19	206.8	17.9	771	13	BJ373284
20	206.6	17.9	858	14	BQ165469
21	206	17.8	914	12	BM815992
22	204.6	17.7	692	14	BF480223
23	204.6	17.7	808	17	BM404603
24	204.4	17.7	955	14	BQ950501
25	204	17.7	738	13	BI423145
26	202.2	17.5	850	12	BF259971
27	202	17.5	631	13	BJ436098
28	202	17.5	731	14	BU004375
29	201	17.4	618	13	BJ433332
30	201	17.4	645	13	BJ430220
31	201	17.4	646	13	BJ433905
32	201	17.4	648	13	BJ430052
33	201	17.4	687	13	BJ435244
34	201	17.4	691	9	AU034101
35	201	17.4	695	13	BJ433499
36	201	17.4	698	13	BJ430606
37	201	17.4	699	13	BJ431424
38	201	17.4	699	13	BJ431971
39	201	17.4	701	13	BJ434216
40	201	17.4	713	13	BJ429678
41	201	17.4	830	14	BQ096649
42	200.4	17.4	610	13	BJ429470
43	200.4	17.4	652	13	BJ373810
44	200.4	17.4	693	13	BJ374845
45	200	17.3	707	13	BJ432258

ALIGNMENTS

RESULT 1
LOCUS AY109333/c
DEFINITION Zea mays CL2757_5 mRNA sequence.
ACCESSION AY109333
VERSION AY109333.1 GI:21212903
KEYWORDS
SOURCE
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitstilt, M.S., Arthur, L.W., Hanarey, M., Morgante, M. and Tingey, S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1912)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
FEATURES
source location/Qualifiers
1..1912
/organism="Zea mays"
/db_xref="taxon:61135"
/db_xref="taxon:4577"
/clone="CL2757_5"

/clone.lib="Maize Mapping Project/Dupont Consensus Library"
 /note-"this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"
 BASE COUNT 361 a 509 c 504 g 390 t 148 others
 ORIGIN

Query Match 23.5%: Score 271.4: DB 11: Length 1912:
 Best Local Similarity 54.8%: Pred. No. 2.1e-71:
 Matches 612: Conservative 0: Mismatches 472: Indels 33: Gaps 4:

2 TGGCAAAACACCTTTTACGTCGAGTCGCTCTGGAAGGCGATCCGTACAATAATGCTG 61
 1601 TCGACACATTCCTCTTACGTCGAGCTGTGTAAGAGGAGACACCTGTACAAGCTGTG 1542
 62 ACCAAATTTGATGCGCTTTAGACGCGATCTTGAACAGATCCGAAGACGCGCTTG 121
 1541 ACCAGCTTCAGATGCTCTTGTGACGCTTGTGAGAGACCTTACAGCAAGCTTG 1482
 122 CTGGCAAACTTACGTAATAAACCGGCGATGTTTGTGCGGCGGAATCCACACAGCG 181
 1481 CTGTGACACCTGACACCAAGACACATGTCATGTGTTGGCGAAGATCAGACCAAG 1422
 182 CTTGGTAGACGCAAGAGATCACCCGTAACCGTTCCGGAATTTGCTATGTGCAAT 241
 1421 CGACCGTGACGACGAGAAATTTGACAGGAGACCTGCGCAACATTTGTTGTGCA 1362
 242 CCGACATGGCTTTGACCTTACCTCTGCGGTTCTACGCGCTATCGGCAACAGCTTC 301
 1361 ACGATCTTGGCTTACACCCACCACTCAAGGTGCTGTGAACATTTAGCAGAGCTCC 1302
 302 CTGACATCAACGAGCGCTTGACCGTGC-----GATCCGCTGGAACAGGCGCGG 352
 1301 CTGATATTGCTCAGGGTGTGATGCGCCTTACCAACGCGCCGAGAGATTTGAGCTG 1242
 353 GTGACAGGCTGTGATCTTTGGCTAACGAACTAATGAACCGACGCTGATGCGCAGAC 412
 1241 GTGACAGGACACATGTTGGGATGCGACGAGACGACCCGCTGAGCTGATGCGCCCTCA 1182
 413 CTATCACTATGACACCGCTGTGTAACGCTGAGGCTGAAGTGTGTAATAACGCGACTC 472
 1181 GCCATGTCTTGCACCAAGCTGGCGCGCTCTCACGAGGTGCGCAAGACGACCT 1122
 473 TGCCGCTGGCGCGCGGAGCGGAAGACGAGTGAATTTTCAGTA-----TG 520
 1121 GCCCTGGCTCAGGCGCGATGGAAGACCGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1062
 521 ACGAGCGCAAAATGTTGATGATGCTGTGCTGCTTTCAGTCAAGACCTCTGAAGAGA 580
 1061 GCCCATGTCCTCCATCGTGTGACACCGCTCTCATCTCCACCGACGACGAGACAG 1002
 581 TCGACCAAAATTCCTGGAAGAGCGGTAATGAAGATCATCAAGCAATTTCTGCCCG 640
 1001 TGACCAAGGATGATGCTCTGACCTGAAGAGACATGTCAAGCTCTCATACCTTG 942
 641 CTGAATGCTGCTCTTCTCCACCAAAATTTCTATCAACCGGCGGCTTTCGTTATCG 700
 941 AGCAATGCTGACGACGAGAAAGACATCTTCAACCTTACCCGCGCGCTTTCGTCATCG 882
 701 GTGCGCAATGCTGACGCTGCTGACTGTGCTAAATTAATCTTGAATCTAGCGCG 760
 881 GCGGCGCCCGGCGATCCGCGCTCACCGCGCGCA-----CAACGACACGTAACGCGG 828
 761 GCATGGCGGCTGACGCTGGGCTGATCTCTGTAAGATCCTCAACAAAGTGAAGCTT 820
 827 GCTGGGAGCCCGACGCGCGCTTCTTCTCGGCAAGAGACCCCAAGAGTGTGCGCA 768
 821 CCGAGCTTACGACGACGATGATGCGGAAACATGCTGTGCTGCGCTGCGCGATC 880

Db 767 GCGGCGCTTACGTCGCGCCAGCAGCGCCCAAGACATCTGTGCGCCNNNNNNNNNNNN 708
 QY 881 GTTGTGAATTCAGGTTTCTTACGCAATCGCGTGGCTGAACGACCTCATGCTAG 940
 Db 707 NNN 648
 QY 941 AACTTGTGATCTGAGAGAAATGCTTCTGAAACATGACCTGCTGTAGTGTCT 1000
 Db 647 ACTGTAGCGGCGCGCGCGATCCCGCAAGGAGATCTTCAATGCTGAGAGACAT 588
 QY 1001 TCGACCTGCGCCATACGCTGATTCATGATGCTGATGCTGCA-----CCGATCT 1054
 Db 587 TCGACTTAGGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 528
 QY 1055 ACAAGAAACCGAGACATACGCTGCTGCTGCTGA 1091
 Db 527 TCATCAAGACCGCGCATACGCGCACTTGTGCGCTGA 491

RESULT 2
 BM807271
 LOCUS
 DEFINITION
 AGENCOURT_6566487 NIH_MGC_88 Homo sapiens cdna clone IMAGE:5738319
 5', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 Mammalia; Eulheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 994)
 NIH-MGC http://mhc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 Plate: L14M12749 row: n column: 16
 High quality sequence stop: 714.
 Location/Qualifiers
 1..994
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5738319"
 /clone.lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: PCMV-SPOB6;
 site_1: NotI; site_2: SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC library."
 BASE COUNT 257 a 198 c 250 g 285 t 4 others
 ORIGIN

Query Match 21.5%: Score 248: DB 14: Length 994:
 Best Local Similarity 58.3%: Pred. No. 2.1e-64:
 Matches 515: Conservative 0: Mismatches 351: Indels 17: Gaps 4:

QY 46 CTTGCAAAATTTGCGCAAAATTTCTGATGCTGCTTACGCGGATCTCGAAGACGAT 105
 Db 2 CCAAGTAAGATTTGCTGACCAATCAGTATGCTGCTTGTGCGGACCTTACAGAGAT 61
 QY 106 CCGAAGACGCGCTTGTGCGAAGCTTACCTAAGACCGGCAATGCTTATGTTGCGCGC 165
 Db 62 CTTGATGCAAGATGCTGTTGAAACATGTTCTAAACATGCAATGATGCTTCTTGTGCGG 121

166 GAAATCAGACAGCGCTGGTAGACATGAGATGATCCCGTAACAGCTTCCGGA 225
 122 GAAATTCATCCAGAGCTGCTGTGACATACCAAAAGTGTGTGACAGCTTTAAAC 181
 226 ATTGGCTATGCTATCCAGATGGCTTTAGACGTAATCTGTGGGCTTTGAGCGCT 285
 182 ATTGATATGATGATTTCTTCCAAAGTTTGTAGTACAAAGACTTGTAACTGTGTAGCC 241
 286 ATGGCAACAGTCTCTGATCATCAACAGCGGCTTGACCTGCC--GATCCGCTGGAA 342
 242 TTGGACCAACAGTCCAGCATTTGTCTCATCTTGTACCAAAATGAAGAAGAC 301
 343 CAGGCGCGGCTGACAGGCTGTGATTTGGCTAGCACTAATGAACGAGCTGCTG 402
 302 ATTGGCTGTGAGACAGCGCTTAATTTTGGCTATGCCATGTAACTGAGAGCTGT 361
 403 ATCCAGACCTATCACTATGACACAGCGTGTGTACAGCGCTGAAGTGGCTAAA 462
 362 ATGCTTTTACCATTTGCTTGGCAGACAAAGTAAATCCAAAGTGGCAGAACTACGCGT 421
 463 AAGGCACTGTGCGGTGCTGCGCCCGGAGCGCAAAAGCGAGTGAATTTCACTATGAC 522
 422 AATGGCACTTGTGCTTGTGACCGCTGATTTAAACTGAAGTACTGTGACGATATG 481
 523 GACGCAAAATCGTTG-----GATCGATGCTGTGCTTTCACACTACAC 570
 482 CAGGATCGAGCTGCTGCTTCCCATCAGAGTCCACACATTTGTATATCTGTACAGAT 541
 571 TGTGAAGATGACAGCAAAATGCTGCAAGAGCGTAATGGAAGATCATCAACCA 630
 542 GATGAAGAGTGTGCTTGTGATGAATGAGGATGCCCTAAAGAGAAAGATCATCAAGCA 601
 631 ATTCTGCGCCCTGAATGCTGACTTCTGCCACAAATTTCTTCATCAACCGGAGCTGCT 690
 602 GTTGTGCTGCGGAATACCTTGATGAGATATCAATCCACTACGCCCACTGTGGAGA 661
 691 TTGCTATCGGTGCGCCCAATGGTGAATGCTGCTGCTGCTGCTGAATTAATTTGCTGAT 750
 662 TTTGTTATTTGTGGGCTCTCAGAGGTGATGCTGTTTGTACTGTGACGAGAAATCATTTGTGAC 721
 751 ACCTACGGCGGATGCGGCTGACAGGTGCGGCTGCTCTGTGTAAAGA-TCCATCAAA 809
 722 ACTTATGCGGCTTGGGCTGCTCATGAGAGAGTGCCTTTTCAGAAAGATTAATACCAAG 781
 810 AGTGAACCTTTCCGCAACCTATGTCGACAGCTATGTCGCAAAACATCTTGTCTGCTG 869
 782 GTTCGACCGTTTCACTGCTCTATGCTGCTGTTGGTGCAAAATCCCTTTTAAAGGAG 841
 870 CCTG-GCCGATGCTGTGAATTCAGATTCTTCTACGCAATCGG 911
 842 TCTGTGCGGAGGAGGTTCTTGTTCANCTCTTATGCTATTGG 884

RESULT 3
 BH770810 920 bp DNA linear GSS 01-MAY-2002
 LOCUS LLMcag552 MG1363 Random Sequence Tag Library Lactococcus lactis
 DEFINITION subsp. cremoris genomic DNA sequence.
 ACCESSION BH770810
 VERSION BH770810.1 GI:20373767
 KEYWORDS GSS.
 SOURCE Lactococcus lactis subsp. cremoris.
 ORGANISM Lactococcus lactis subsp. cremoris.
 Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
 Lactococcus.
 1 (bases 1 to 920)
 REFERENCE
 AUTHORS Bolojin, A., Ehrlich, S.D. and Sorokin, A.
 TITLE Studies of genomes of dairy bacteria Lactococcus lactis
 JOURNAL Sci. Alimentis, (2002) in press
 COMMENT Contact: Sorokin A
 Genetique Microbiologie
 INRA

CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
 Tel: 33 1 34 65 25 16
 Fax: 33 1 34 65 25 21
 Email: sorokine@jouy.inra.fr
 best homologue in strain IL1403 is metK (908)
 Class: shotgun
 High quality sequence start: 30
 High quality sequence stop: 893.
 Location/Qualifiers
 source 1..920
 /organism="Lactococcus lactis subsp. cremoris"
 /strain="MG1363"
 /db_xref="taxon:1359"
 /clone_lib="MG1363 Random Sequence Tag Library"
 /note="Vector: pSGWU2. Site_1: Small library of
 chromosomal fragments of L.lactis strain MG1363 was
 prepared by partial AluI digestion or by sonication."

BASE COUNT 271 a 169 c 207 g 273 t
 ORIGIN

Query Match 21.4% Score 246.6; DB 17; Length 920:
 Best Local Similarity 59.1%; Pred. No. 5,3e-64;
 Matches 482; Conservative 0; Mismatches 319; Indels 14; Gaps 3:

346 GCGCGCGTGACAGGCTGTGATTTGGCTACGCACTAATGAACGAGCTGTGATG 405
 60 GGTGTGCTGACCAAGAAATGATGCTGTTTGCACATACGAAATGCCAATACATG 119
 406 CCAGACCTATTCACCTATGACACAGCTGTGTACAGCGCTGAGCTGAAGCTAAAC 465
 120 CCACCTTGCCATTTTCACTTTCACATTAATTAAGAAATTTGCTGATTTGCTTAATCA 179
 466 GGCACCTGCGCGGCTGCGCGCGGAGCGGAAAGCAGCTGATTTTACATG--AC 522
 180 GGTGAATTTTCACTTCTGCTCCAGATGCAAAATCAACAGTACTGATGATTAAT 239
 523 GACGCAAAATCGTGTATGATGCTGTGCTTTCACATGACGACTGTAAGATC 582
 240 AATGGAAGCAAAAGCTGTGTATACCGGTATCTGACATGACATGCTGTACGGCA 299
 583 GACCAGAAATCGTGCAGAGAGCGGTATGAAAGATCATCAACCAATTTGCCCT 642
 300 ACAAAAGAAATTCATGATGATGATTAACAAAGTGTATTAAGCTATTAATGCTGCC 359
 643 GAATGCTGACTTGTGCGACCAATTCCTGATCAACCGCGGTGCTTGTATGCGT 702
 360 GAATTTTGGATGATGAACATATATCTTATTAATCAACATGACGTTTGTATGCGT 419
 703 GGCCCAATGGGTGACTGCGCTGCTGACTGTGCTGAATTAATTCGTATACCTACGCGGC 762
 420 GTTCCTCAAGAGACACAGGTTTGACTGGCGGCAAGATTATTTGTATACAGGGGGG 479
 763 ATGGCGGCTACGCGTGGCT-----GCATTTCTGTGTAAGATCCATAAAGTGACC 817
 480 GTTATGCACCATGATGATGTTGGGCGCTTCTCAGGTAAGATGCAACTAAAGTTGACC 539
 818 GTTCCCGACGCTACGACGACCTATGTCGCGGAAAAACATCGTTGCTGCGGCTGCGG 877
 540 GTTCACCTCTTACCGGCGCTGCTTATGTGCAAAACATTTGCGCTGCGGACTTGCTG 599
 878 ATCGTTGTGAATTCAGGTTTCTTACGCAATCGCGCTGAGTGAACGACCTCCATATG 937
 600 ATAAGCTCAAAATTCACATTTCTTATGCGATGTGTGCGCACTGCTACATCATCAATG 639
 938 TAGAACTTTGGTACTGAGAAAGCTCTTGAACACTGACCTGTGTGATCGTATG 997
 660 TTGACACTTTTGGTAGAGTAAGTTTCAATATGATGATGATGATGATGATGATGATG 719
 998 TCTTGCACCTGCGCCCATACGCTGATGATGATGATGATGATGATGATGATGATGATG 1057
 720 TCTATCTACTCCGTGATACGAAATTAATGAGGCTTGTGACGCTTGTATTAATCATG 779
 1058 AAGAAACCGAGCATACGCTTGTGCTGTAACAT-----TTCGCTGGGAAAAA 1111

[illegible]

	Query Match	20.3%	Score	235	DB	17	Length	899	
	Best Local Similarity	54.7%	Pred.	No.1	ge-60:				
	Matches	481	Conservative	16	Mismatches	388	Indels	15	Gaps
									2:
OY	241	TCCGACATAGGGCTTGTGAACGCTAACTCCTATCGGTTTCGACCGCATCGCAACACTCT	300						
Db	10	TTCCGATAAGGGTTTTGCATTACAGACCCTTAACGTTGGTTGGTGGAACAACAGTCT	69						
OY	301	CCTGACATCAACAGAGCGGTTGACCGGTGCATCCGCTGGAAACAG--GGCGCGGTGAC	357						
Db	70	CCAACATTTGCTCAGAGGTTTTGCATTACACMACAGCGTAGAAGACTAGTGGCGGTGAC	129						
OY	358	CAGGTCGTGATGTTGGCTAGCACACTAATAAACCAGCGTGTGATGCCAGCACTATC	417						
Db	130	CAAGGATATCATGTTTGGATTASGCCAGCAGACCAAASCAGAAAGTTGCTTACACTT	189						
OY	418	ACCTATGCACACCCCTCTGTGTGACAGCGTACGCTCAAGTCCGTAAAAAACGGCATCTCCG	477						
Db	190	TTGTTGGCCCAACAGCTATCAACATGGCCATGGCCCAACGCCAAGAAGACGGCTTTTGS	249						
OY	478	TGCGTGGCCCCGAGACGGCAAAAGCCAGTGACTTTTCAGTATGACGACG-----	526						
Db	250	TGGTTGAAASSSTGASASACMASSSAGGTGACCGTSGAGTCAAGAGAAAGACASAGTGA	309						
OY	527	-GCAAAATCCGTTGCTATCGATGCTGTGCTGCTTCCATCCAGCACTGTGAAGAGATGAC	585						
Db	310	TGGAATSCATTGAGAAATTTGACACCGCTGTGTGTGTGCGCAACAGCGCCAGACAGTCTCC	369						
OY	586	CAGAAATCGCTGCAAGAAGCGGTAAATGGAAGATATCATCAAGCCAAATTTCCCGCTGA	645						
Db	370	ACCGAAGACTTGAAGATCTTTGATCAAGTCGCAATCATCAAAAGTATCTBACAGCCGAC	429						
OY	646	TGGCTACTTCTGCCACCAATTTCTTCATCAACCCGACGGGTGTTTCGTTATCGTGCC	705						
Db	430	ATGTTGGAGCAAAAGAACCAAGTACTACATCCAGCCTTCCGCAATTCGTATTGGTGTG	489						
OY	706	CCAATGGGTGACTGGGCTCTGACTGGTGTAAATTAATCGTTGTAATCTAGGGGGCATG	765						
Db	490	CCTCAAGGTGAAGCGGGTTTGAACGGTGAAGAATCATCTGTCGACGGCTTAGCGGTGTC	549						
OY	766	GCGGCTCAAGGTTGGCGGTGTGATCTCTGTGAAGATCCATCAAAAGTGGACGTTCCGA	825						
Db	550	TCTGCGGTGGTGGTGGTGGCTCTTCCGGTAAAGACATCTCAAGGTGGAACGTTCCGCT	609						
OY	826	GCTTACGACGACGTTATGTGCGCAAAACATCTGTTGCTGTGGCTGTGGCCGATCGTTGT	885						
Db	610	GCTCTGCGCGCAGATGGGTGGGCAAGTCCTTGGGTGACSCSCSGTTTSTSCAAGAGATC	669						
OY	886	GAAATTCAGGTTTCTTACGCCATGGCGCTGGCTGAACCGACACTCCATATATGTTAGAACT	945						
Db	670	CAAGGCAATTCCTCTACGCGATCGGTATTTCCCGAGCCTCTGTCTTCCACGTGCACAC	729						
OY	946	TTTGGTACTGAGAAAGTCCCTTCTTAACAACATGACCCCTGTGTGCTGAGTTCTTCGAC	1005						
Db	730	TACGTTACCCGCCACCAAGTCCGACGAGGAATAATCATCGATATCATAAAGAATCTTAC	789						
OY	1006	CTGCGCCCATACGGTCTGTATTCAGATGCTGATGCTGACACCCCATCTACAAAGAAC	1065						
Db	790	TTGAAACCAAGGTGTGTGTGTAAGAAGTTGGACTTGGCTAGACCAATCTACTTGCACACT	849						
OY	1066	GCAGCATACGCTACTTGTGTCGTGAACATTTCCCCTGGG	1105						
Db	850	GCTTCTACGCGTCACTTTACCAACCAASAATVCCCATGGG	889						
RESULT 5									
LOCUS	BE035235	1056 bp	mRNA						
DEFINITION	MO02F06. NO Mesembryanthemum crystallinum cDNA 5' similar to s-adenosylmethionine synthetase #, mRNA sequence.								
ACCESSION	BE035235								
VERSION	BE035235.1	GI:8330359							
KEYWORDS	EST.								


```

RESULT 9
LOCUS      B0055329
DEFINITION AGENCOURT_6797446 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5807491
ACCESSION  B0055329
VERSION     B0055329.1 GI:19814669
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1071)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/
TITLE       Unpublished (1999)
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: Lou Staudt
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LCM2049 row: p column: 20
            High quality sequence stop: 714.
            Location/Qualifiers
                1..1071
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5807491"
                /clone_id="NIH_MGC_99"
                /tissue_type="lymphoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: lymph. Vector: pOT87. Site.1: XhoI; Site.2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(c). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the Laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-CDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                Library."
BASE COUNT  279 a 238 c 263 g 287 t 4 others
ORIGIN
Query Match 19.08: Score 219.2; DB 14; Length 1071;
Best Local Similarity: 57.98; Pred. No. 1.4e-55;
Matches 453; Conservative 0; Mismatches 314; Indels 16; Gaps 3;
QY 3 GCGAAGACCGCTTTTACGTCGCGATCGCTCTGAAGGCGATCCTGACAAATTGCTGA 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 29 GGGCAGCATCTCTTTCACCTCAGAGTCGTCGGGAGGACCCAGCATTAAGATTGTA 88
QY 63 CCAATTTCTATGCGCTTTAGACGCATCTCGAAGAGATCCGAAAGCAGCGCTGC 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 89 CCAATTCAGTATGCTGCTTGAATGCCATCTGATGCCAAAGTAGC 148
QY 123 TTGCAACCTAAGTAAAGCGGCGATTTTATGTCGGCGGAAATGACACAGCGC 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 149 TTGCAAACTGTTCTAATAATCGAATGATCTTCTGCGGAAATTAATACATCAGAGC 208
QY 183 CTGGGTACATCGAAGAGATCACCGGTACACCGTTGCGAAATTTGCTATGTCATTG 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 TGTCTTGTACACGAGAGTGTGTCGAGAGCTTTAAACATTCGATGATGATTTC 268
QY 243 CGACATGGCGCTTGAACGCTACTCTGTGCGGTTCTGAGCCCTATCGGCAACAGTCTCC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 TTCCAAAGCTTTTACACTACAGACTTGTACGCTGTTGACCTTTGAGCAAGTCTACC 328
QY 303 TGACATCAACGAGCGGTGACCGTGCC---GATCCGCTGGAACAGCGCGGGTGACCA 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 329 AGATATTGCTCAGAGTGTTCATCTTGACAGAAATGAAGAGACATTGCTGTGAGACCA 388
QY 360 GGGTCTGATGTTTGGCTACACCACTAATGAACCGAGCTGTGATGCCAGCCTATCAC 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 389 GGGCTTATGTTTGGCTATGCCATGATGCAAGTGAAGAGAGTGTATGCTTTAAACATTGT 448
QY 420 CTATGCACACCGCTGTGTAGACGCGTCAAGAGTGGTAAAGGCGACTTCGCGTG 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 449 CTGGCACACAGCTAAATGCCAAACTGSCAGAACTACGCGCTAATGGCACTTGTGCTTG 508
QY 480 GCTGCGCCGCGAGCGGAAAGCCAGCTGCTCTTTCAGTATGACGACGCAAAATCGTTTC- 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 509 GTTACGCCCTGATTTCTAATAACTCAAGTTACTGTGACGATATGAGAGATCGAGGTGCTG 568
QY 539 -----CTATCGATGCTGTGCTGCTTTCCACTCAGACACTCTGAGAGATCGACA 587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 569 GCTTCCCATCAGATCCACACCAATTTGTAATCTGTTCCATGATGACAGAGCTTTGCT 628
QY 588 GAATTCGCTGCAAGACCGGTAAATGGAAGATATCATCAAGCAATTCGCGCGCTGAATG 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 629 TGATGAATGAGGATGCCCTAAAGAGAAAGTATCATAAAGCACTGTGCTCGGAAATA 688
QY 648 GCTGACTTCTGCCACCAAAATCTTCATCAACCCGAGCGCTGTTGCTTATCGGTGCC 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 689 CTTGATGATGATGATCAATCTACACCTACAGCCAAAGTGGCGATTTGTTATGGTGCGCC 748
QY 708 AATGCGTACGTCGCGTCTGACTGCTGTAATTAATGCTGATACCTACAGCGCGCATGCC 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 749 TCAGCTGATGCTGCTGTTGACTGAGCG-CAATCATTTGTGACACTATGCGCGTTGGG 807
QY 768 GCG 770
DB 808 GTG 810

```

RESULT 10
LOCUS B0964578
DEFINITION AGENCOURT_10050435 NIH_MGC_134 Mus musculus cDNA clone
ACCESSION B0964578
VERSION B0964578.1 GI:22380056
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 977)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM14089 row: p column: 05
 High quality sequence start: 213
 High quality sequence stop: 405.
 Location/Qualifiers

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
LOCUS B1921837	B1921837	B1921837	B1921837	B1921837	Tomato	Tomato
DEFINITION	EST5741740 tomato callus Lycopersicon esculentum cDNA clone					
ACCESSION	B1921837					
VERSION	B1921837.1					
KEYWORDS	EST					
ORGANISM	Lycopersicon esculentum					
SOURCE	Tomato					
REFERENCE	Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai, J., Utterback, T., Van Aken, S., Romling, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.					
AUTHORS	Generation of ESTs from tomato callus tissue (2001)					
TITLE	Unpublished (2001)					
JOURNAL						

COMMENT	Contact: CUGI Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html This clone is available through the Clemson University Genomics Institute Seq primer: T3.
FEATURES	Location/Qualifiers
SOURCE	1. :755 /organism="Lycopersicon esculentum" /cultivar="TA96" /db_xref="taxon:4081" /clone="CLEC7AD16" /clone_1lb="tomato callus" /cluse_type="callus" /dev_stage="25-40 days old" /lab_host="XLI-Blue MRF" /note="vector: pBluescript SK(+) Site_1: EcoRI; Site_2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research; clec - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"
BASE COUNT	210 a 196 c 184 g 165 t
ORIGIN	
Query Match	18.1% Score 209; DB 13; Length 755;
Best Local Similarity	57.3% Pred. No. 1.5e-52;
Matches 430:	Conservative 0; Mismatches 300; Indels 21; Gaps 2;
OY	37 GAAGGCGATCCTGCACAAATTGGTCGACCATAATTTGTGATGCCGTTTAGACGCGATCC 96
DB	2 GAAGGTCAKCCCAGCACACTCTGTGACCAAGCTCATGCTCATCTTGTGATGCTTGCTG 61
OY	97 GAACAGATCCGAAGCAGCGCTTCTCTTGGCAAACCTACGTAAACCCGGCATGGTTTA 156
DB	62 GAGCAGGATCCGGAAGAAGAGGTTCATGTGAACCTGCACAAAGACAACATGGTTATG 121
OY	157 GTTGGCGGCAATCAACACAGCGCCTGTGGTAGACATCGAAGAGATGACCCGTAAACC 216
DB	122 GCTCTTGGAGAGATGCACAAACGAGCCACACCGTAGACTACAGAACATCGTACGTAC 181
OY	217 GTTCCGGAATTGCTATGTGCAATTCGACATGAGGCTTTGACGCTAACTCCTGTGGGTT 276
DB	182 TGCAAGGCGATGGGTGCTGCTCAGACAGATGTTGGTCTTGATGTCGACAACTGCAANG 241
OY	277 CTGACGCGTATGCGCAACAGAGTCTCTGATCATCAACACAGGCGTTGACCGT----- 327
DB	242 CTGTGTAACATTTGAAGAACAGAGCCCTCATGCATCGCCCAAGGATTACAGCGTATCTT 301
OY	328 GCCGATCCGCTGGAACAGAGCGCGGGGTACACCGGCTGTGATTTTGGCTAGCACAATAT 387
DB	302 AAGAAACCGAAGAAATTGGAGCTGTGTACCAAGTGCATATGTTTGGGTATGCCACAGAT 361
OY	388 GAAACCGAGTGCATGATGCGACAGCACACTATCATCTATGTCACACCGTCTGTGTACAGCGT 447
DB	362 GAAACTCGTGCAGCTCATGGCCCCCTCACCCATGTTTTGGCGACACCAACACTTGTGTCCAAGCTT 421
OY	448 GCTGAAGTGCCTAAAGACGCGACCTCTGCCGCTGCTGCGCCCGGACGCGCAAAAGCCAGGTG 507
DB	422 ACTGGAAGTAGAAGAAACCAACCTGCCATGGCTCAGGCCGGAATGCGCAAGACCCAGATT 481
OY	508 ACTTTTCAGTATGACGAGCGGCAAAATCG-----TTGGTATGATGATCCTCTGTG 555
DB	482 ACTGTTGATGATCAAAAGACGACCAATGTGTGCATGTGCCCTATTATGAGTTTCACACTGTTTC 541
OY	556 CTTTCCACGACGACTGTGAAGAGATGCGACGAGAAATCGCTGCGCAAGAGCGGTAATGNA 615
DB	542 ATCTCAACCCAGCATGACGAACCTGTCAACACGACAGATGCGCCAGGACCTTGAANAG 601
OY	616 GAGATCATCAAGCAATTCCTGCCGCTGAATGCTGACTTCTGCACCAAAATTTCTTCATC 675

Db 602 CATGTGATCAAGCCTGTGATCCAGCTAAGTACCTGACAGAAACACCATCTTCACCTC 661
 Oy 676 AACCCGACCGTCTGCTTCTTATCGGTGGCCATGGGTGACTCGGTGACGTGCT 735
 Db 662 AACCCATCAGTCCCTGCTTGTGATGTCCTCCACATGAGTGTGCTGCTTACTGAGG 721
 Oy 735 AAATATCGTGTACCTACCGCCGATCG 766
 Db 722 AAATCATCATGTGATCTACCTACGAGAGCTGG 752
 RESULT 13
 BG414252 817 bp mRNA linear EST 23-Oct-2001
 LOCUS HXSMK0002012f Hordeum vulgare testa/pericarp EST library
 DEFINITION HVCDNA0013 (normal) Hordeum vulgare cDNA clone HXSMK0002012f, mRNA
 sequence.
 ACCESSION BG414252
 VERSION BG414252.1 GI:13319899
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 REFERENCE Hordeum vulgare: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 : Triticeae; Hordeum.
 TITLE 1 (bases 1 to 817)
 JOURNAL Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Kannangara, G., von
 Weizstein, D., Akhunov, E., Chin, A., Choi, D.W., Fenton, R.D., Kianian
 , P., Otto, C., Simons, K., Zhang, D., Begum, D., Fritsch, D., Yu, Y.,
 Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex testa/pericarp cDNA library
 unpublished (2001)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Total hg bases - 347
 Seq primer: AATTACCTCCTCACTAAGG
 High quality sequence stop: 643.
 FEATURES
 source
 1. 817
 Location/Qualifiers
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HXSMK0002012f"
 /clone_1lb="Hordeum vulgare testa/pericarp EST library
 HVCDNA0013 (normal)"
 /tissue_type="testa/pericarp"
 /lab_host="TJC121"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 plants were raised from seeds in a Controlled Environments
 growth chamber maintained in continuous light at 18°C, and
 testa and pericarp were dissected from developing kernels
 at Washington State University, Pullman, WA (Kannangara,
 von Weizstein). Total RNA was prepared, poly(A) RNA was
 purified, one cDNA library was made, and 1 million ptu
 were in vivo excised to give Bluescript SK(-) cDNA
 phagemids in the TJC121 lab at the University of
 California, Riverside (Akunov, Chin, Choi, Close, Fenton,
 Kianian, Otto, Simons, Zhang). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Fritsch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons, Oates
 , Rambo, Main). The sequence has been trimmed to remove
 vector sequence and contains a minimum of 100 bases of
 phred value 20 or above. For more details on library
 preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders
 see close 'TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"
 BASE COUNT 182 a 256 c 219 g 160 t
 ORIGIN
 Query Match 18.0% Score 207.6; DB 12; Length 817;
 Best Local Similarity 56.8%; Pred. No. 4.3e-52;
 Matches 434; Conservative 0; Mismatches 309; Indels 21; Gaps 2;
 Oy 12 CCTTTTACGTCGAGTCCGCTCTCTGAAAGGACCTGACAAATTCCTGACAAATTC 71
 Db 53 CCTTTTACGTCGAGTCCGCTCTCTGAAAGGACCTGACAAATTCCTGACAAATTC 112
 Oy 72 TGATGCCGTTTACAGCCGATCCTGCAACAGATCCGAAGACGCTTCTGCGAAG 131
 Db 113 CGAGCGCGTCTGAGACCCCTGCTGGCCCAAGATCTGACAGCAAGGTTGCTCGAGAC 172
 Oy 132 CTAGTAAACCCGATGTTTACGTTGGCGGAATGACACACGCGCTGGTGA 191
 Db 173 CTGACCAAGACCAACATGTCATGCTTCTGAGATACGCAACGACACCGTGA 232
 Oy 192 CATTGAGAGATCACCCGTAACACCGTTCGCAAAATTCGCTATGCTTCGACATGG 251
 Db 233 CTATGAGAGATCGTGTGACACCTGCGCAATTCGCTATGCTTCGACATGG 292
 Oy 252 CTTTACCGCTTACCTCTGTCGCGTTCTGAGCGCTATGCGCAACAGTCTCTGACATCA 311
 Db 293 TCTTCGACCGCGACCATTCGCAAGGTGCTGTCACATGAGACCAACATCCCTGACATTC 352
 Oy 312 CCAGGCGGTTGACCGTGC-----GATCCGCTGAACAGGCGGCGTGCAGGG 362
 Db 353 CCAGGCGGTTGACCGTGCACATGTCACCAAGCTTCCGAAGAGTGCAGGCGTGCAGGG 412
 Oy 363 TCTGATGTTTGGCTACGCACTAATGAACGAGCTGCTGATCCACACCTATCAGCTA 422
 Db 413 CATCATGTTTGGCTATGCAACGAGTGAAGACCCCTGACATGATGCTTCCACACCATGCT 472
 Oy 423 TGCAACCGCTGCTGACAGCTGACAGTGAAGTCCGTAACACGCACTCTGCGTGGCT 482
 Db 473 TGCCACCAAGCTCGAGCTGCGCTACACCGAGGTCCGACAGATGCACTCTGCTGCTGCT 532
 Oy 483 GCGCGCGGAGCGGAAGCCAGGTCGACTTTTCACTA-----TGACGAGGCGAA 530
 Db 533 GAGGCTGATGACAGACCAAGCTCACCATTGAGTACCTAATGAGGTGCTGCTATGCT 592
 Oy 531 AATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 590
 Db 593 GCGTGTGCTGTGCAACCGCTCTCATCTTACCCAGATGACAGACGCTACCAACGA 652
 Oy 591 ATGCTGCAAGACGCTAATGAGAGATCATCAACCAATTCCTGCTGCTGCTGCTGCT 650
 Db 653 TGAATGCGCGCGGACCTCAAGAGAGACGCTTCAACCGCTGATTCGCGGAGTACCT 712
 Oy 651 GACTTTCGCAACCAATTCATCAACCGGAGGCTGCTTCTGCTTACCGTGGCCCAAT 710
 Db 713 CGATGAACCAATTCATCAACCGGAGGCTGCTTCTGCTTACCGTGGCCCAAT 772
 Oy 711 GGGTGAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 754
 Db 773 TGGGATGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
 RESULT 14
 BJ432712 720 bp mRNA linear EST 13-MAR-2002
 LOCUS BJ432712 Dictyostelium discoideum cDNA library, VF Dictyostelium
 DEFINITION BJ432712 Dictyostelium discoideum cDNA clone d419c17 3', mRNA sequence.
 ACCESSION BJ432712
 VERSION BJ432712.1 GI:19407434

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:03:55 ; Search time 3355.45 Seconds

(without alignments)
10459.998 Million cell updates/sec

Title: US-09-622-419-3

Perfect score: 1206

Sequence: 1 atgaacgttttaaccgccg.....tggaattatgttgatataa 1206

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_in:*
- 3: gb_in:*
- 4: gb_in:*
- 5: gb_in:*
- 6: gb_in:*
- 7: gb_in:*
- 8: gb_in:*
- 9: gb_in:*
- 10: gb_in:*
- 11: gb_in:*
- 12: gb_in:*
- 13: gb_in:*
- 14: gb_in:*
- 15: gb_in:*
- 16: gb_in:*
- 17: gb_in:*
- 18: gb_in:*
- 19: gb_in:*
- 20: gb_in:*
- 21: gb_in:*
- 22: gb_in:*
- 23: gb_in:*
- 24: gb_in:*
- 25: gb_in:*
- 26: gb_in:*
- 27: gb_in:*
- 28: gb_in:*
- 29: gb_in:*
- 30: gb_in:*
- 31: gb_in:*
- 32: gb_in:*
- 33: gb_in:*
- 34: gb_in:*
- 35: gb_in:*
- 36: gb_in:*
- 37: gb_in:*
- 38: gb_in:*
- 39: gb_in:*
- 40: gb_in:*
- 41: gb_in:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1206	100.0	1206	AB055109	AB055109 Escherich
2	1206	100.0	1206	AX018962	AX018962 Sequence
3	1206	100.0	3465	AX000478	AX000478 Sequence
4	1206	100.0	3794	AX000474	AX000474 Sequence
5	1206	100.0	3794	AX018970	AX018970 Sequence
6	1206	100.0	4975	AX018972	AX018972 Sequence
7	1206	100.0	12058	AE000364	AE000364 Escherich
8	1206	100.0	71128	ECU29581	ECU29581 Escherich
9	1205	99.9	1216	AX000470	AX000470 Sequence
10	1167.6	96.8	26658	AP002563	AP002563 Escherich
11	1166.8	96.7	6488	AE005509	AE005509 Escherich
12	939	77.9	251891	AC084742	AC084742 Mus muscu
13	833.2	69.1	20653	AE008836	AE008836 Salmonell
14	823.6	68.3	230050	AL627277	AL627277 Salmonell
15	514.8	42.7	9280	AE013916	AE013916 Yersinia
16	514.8	42.7	210050	AJ414146	AJ414146 Yersinia
17	443	36.7	11773	AC023612	AC023612 Mus muscu
18	335.4	27.8	10786	AE004301	AE004301 Vibrio ch
19	268.6	22.3	2816	AB024603	AB024603 Burkholder
20	268.6	22.3	12537	AB024602	AB024602 Pseudomon
21	259.4	21.5	6866	ECB301654	ECB301654 Erwinia c
22	258.6	21.4	347660	AP002994	AP002994 Mesorhizo
23	257.2	21.3	10235	AE004786	AE004786 Pseudomon
24	252	20.9	11674	AE005391	AE005391 Escherich
25	252	20.9	257071	AE002558	AE002558 Escherich
26	251	20.8	217420	BSUB0017	BSUB0017 Bacillus su
27	244	20.2	1221	AB055108	AB055108 Escherich
28	244	20.2	1221	AX018966	AX018966 Sequence
29	244	20.2	11684	AE000263	AE000263 Escherich
30	244	20.2	13823	D90812	D90812 E. coli geno
31	244	20.2	19521	D90811	D90811 E. coli geno
32	242.2	20.1	14807	AB024601	AB024601 Pseudomon
33	237.8	19.7	13391	AE012389	AE012389 Xanthomon
34	237.2	19.7	10530	AE013796	AE013796 Yersinia
35	234.6	19.5	1217	AX436324	AX436324 Sequence
36	234.6	19.5	1224	AX436323	AX436323 Sequence
37	234	19.4	11137	AE005860	AE005860 Caulobact
38	225	18.7	4160	AX416825	AX416825 Sequence
39	224.6	18.6	14857	AE009543	AE009543 Brucella
40	221.2	18.3	40281	MLCB22	298741 Mycobacteri
41	221.2	18.3	348450	MLCPRN4	AL583920 Mycobacte
42	220.8	18.3	15057	AE003977	AE003977 Yersinia f
43	217.4	18.0	11206	AE009136	AE009136 Agrobacte
44	217.4	18.0	12426	AE008102	AE008102 Agrobacte
45	217	18.0	9988	AE011935	AE011935 Xanthomon

ALIGNMENTS

RESULT 1
AB055109
LOCUS
DEFINITION AB055109 1206 bp DNA linear BCT 14-APR-2001
desulfinate, complete cos.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Escherichia coli (strain:JM109) DNA.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia coli
1 (sites)
REFERENCE Minara,H., Kurihara,T., Yoshimura,T., Sode,K. and Esaki,N.
AUTHORS Cysteine sulfinate desulfinate, a NIFS-like protein of Escherichia

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE

|||||
Db 1201 GATTAA 1206

RESULT 2
LOCUS AX018962
DEFINITION Sequence 3 from Patent WO942591.
ACCESSION AX018962
VERSION AX018962.1 GI:10043057
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia coli.

REFERENCE 1 (bases 1 to 1206)
AUTHORS Schroeder, H.
TITLE Method for producing biotin
JOURNAL Patent: WO 942591-A 3 26-AUG-1999;
SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES
Location/Qualifiers
1..1206
/organism="Escherichia coli"
/db_xref="taxon:562"
1..1206
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07715.1"
/db_xref="GI:10043058"
/translation="MVNENPAQFRAQFPALODAGVYLDLSAATALKPEAVVEATQOFTS
LSAGNVHRSOPFAEORBLAREEAAREKVAQOLNAPDDKTIWTRGTSTIMVAVOCVA
RPRLOPGEITIVSAEHNANLVPILNVAQOLNAPDDKTIWTRGTSTIMVAVOCVA
SRILALGMSVNTGCPDLARATIPAHAGVAVVNDGQGVNHPADVQOIIDIDYAT
SGHRLYGTGIVLYGKSELLEAMSPLWGGKRMVHEVSFDFTTOSAPKLEAGTPNV
AGVIGLSALEMLADYDINDQAESWSRSRLATLAEDALARPFRSFRCDSSLEAFDA
GVHSDWYTLLEAVGIALRACQHCQAPLAEELVGTILRASFAFYPTKSDVDALNVAV
DRALELVLD"

CDS

BASE COUNT 256 a 315 c 359 g 276 t
ORIGIN
Query Match 100.0%; Score 1206; DB 6; Length 1206;
Best Local Similarity 100.0%; Pred. No. 1e-301;
Matches 1206: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ATGAACGTTTAAATCCCGCGAGTTTCCGCCACTTACAGATGCGGCG 60
Db 1 ATGAACGTTTAAATCCCGCGAGTTTCCGCCACTTACAGATGCGGCG 60
Oy 61 GCTATCTGACAGCGCGCGAGTTTAAACCTGAAGCGGTGTTAGCCACCA 120
Db 61 GCTATCTGACAGCGCGCGAGTTTAAACCTGAAGCGGTGTTAGCCACCA 120
Oy 121 CAGTTTACAGTCTGAGCGCGCGAAGCTCATCGAGCGAGTTTGCAGAGCCCAACGC 180
Db 121 CAGTTTACAGTCTGAGCGCGCGAAGCTCATCGAGCGAGTTTGCAGAGCCCAACGC 180
Oy 181 CTGACCGCGCGTTATGACCTGCACAGAGAAAGTGGCCAAATTACTGATGACCGGAT 240
Db 181 CTGACCGCGCGTTATGACCTGCACAGAGAAAGTGGCCAAATTACTGATGACCGGAT 240
Oy 241 GATAAACATATGCTGAGAGCGCGCGACACTGATCATCAACTTGGTGGCACAATGC 300
Db 241 GATAAACATATGCTGAGAGCGCGCGACACTGATCATCAACTTGGTGGCACAATGC 300
Oy 301 TATGCGCGTCCGCTGACACCGCGCGATGATATTGTGACGTTGCGACAGACAC 360
Db 301 TATGCGCGTCCGCTGACACCGCGCGATGATATTGTGACGTTGCGACAGACAC 360
Oy 361 GCCAACCTGTCCTGCTGATGCTGCGCCAAACAACCTGAGACCAACTGCTGAATTG 420
Db 361 GCCAACCTGTCCTGCTGATGCTGCGCCAAACAACCTGAGACCAACTGCTGAATTG 420

Oy 421 CCGCTTAATGCGCAGCAGCTGCGGATGTCGATTTGTTGCCAGAACTGATTAATCCCGT 480
Db 421 CCGCTTAATGCGCAGCAGCTGCGGATGTCGATTTGTTGCCAGAACTGATTAATCCCGT 480
Oy 481 AGTCGGATTTGCGGCTGGGTCAGATGTCGAACGTTACTGGCGGTTCCCGGATTCGCG 540
Db 481 AGTCGGATTTGCGGCTGGGTCAGATGTCGAACGTTACTGGCGGTTCCCGGATTCGCG 540
Oy 541 CGAGCGATTACCTTTGCTCATTACGCGGAGATGGTGTGATGGTGTGATGGTGTGAGCG 600
Db 541 CGAGCGATTACCTTTGCTCATTACGCGGAGATGGTGTGATGGTGTGATGGTGTGAGCG 600
Oy 601 CGAGTGCATTTCCCGCGGATGTTGACAACTGATTAATGATTTGATTTTTCAGGT 660
Db 601 CGAGTGCATTTCCCGCGGATGTTGACAACTGATTAATGATTTGATTTTTCAGGT 660
Oy 661 CACAACCTGTATGCGCGCAGACAGTATGCGGCTGCTATGTAATTCAGAACTGCTGAG 720
Db 661 CACAACCTGTATGCGCGCAGACAGTATGCGGCTGCTATGTAATTCAGAACTGCTGAG 720
Oy 721 CGGATGTCGCGCTGGCTGGCGCGGCAAAATGCTTCAAGATGATTTGACCGCTTC 780
Db 721 CGGATGTCGCGCTGGCTGGCGCGGCAAAATGCTTCAAGATGATTTGACCGCTTC 780
Oy 781 ACGACTCAATCTGCGCGCGGAGAACTGGAAGCTGGAACGCCAAATGCTGCTGCATA 840
Db 781 ACGACTCAATCTGCGCGCGGAGAACTGGAAGCTGGAACGCCAAATGCTGCTGCATA 840
Oy 841 GGATTAAGCGCGCGCTGGAATGCTGCGGAGATTCAGATATACACAGCGCGCAAGCTGC 900
Db 841 GGATTAAGCGCGCGCTGGAATGCTGCGGAGATTCAGATATACACAGCGCGCAAGCTGC 900
Oy 901 ACCGCTAGCTTGAACAGCTGCGGAGAGATGCGCTGGCGAAGCTGCCGCTTTCCTCA 960
Db 901 ACCGCTAGCTTGAACAGCTGCGGAGAGATGCGCTGGCGAAGCTGCCGCTTTCCTCA 960
Oy 961 TTCGCTGCGAGATTTCCAGCTGCTGCGCTTGAATTTGCTGCGGCTTCATCATAGCAT 1020
Db 961 TTCGCTGCGAGATTTCCAGCTGCTGCGCTTGAATTTGCTGCGGCTTCATCATAGCAT 1020
Oy 1021 ATGTGACGCTCTGCGCGAGTACGATGATGCTGCGGCGCGGAGCATTCGCTCAG 1080
Db 1021 ATGTGACGCTCTGCGCGAGTACGATGATGCTGCGGCGCGGAGCATTCGCTCAG 1080
Oy 1081 CCGTACTGCGAGATTAAGCGCTAACCGGACACACTGCGGCTCTTTTGGCCATATAT 1140
Db 1141 CCGTACTGCGAGATTAAGCGCTAACCGGACACACTGCGGCTCTTTTGGCCATATAT 1140
Oy 1141 ACAAGAGTATGATGATGCTGCTGCTGATGATGCTTGAACCGCGCGCTGGAATTAATG 1200
Db 1141 ACAAGAGTATGATGATGCTGCTGCTGATGATGCTTGAACCGCGCGCTGGAATTAATG 1200
Oy 1201 GATTAA 1206
Db 1201 GATTAA 1206

RESULT 3
LOCUS AX000478
DEFINITION Sequence 9 from Patent WO9905285.
ACCESSION AX000478
VERSION AX000478.1 GI:7240887
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3465)
AUTHORS Schroeder, H. and Hauer, B.
TITLE METHOD FOR PRODUCING BIOTIN
JOURNAL Patent: WO 9905285-A 9 04-FEB-1999;
SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES
Location/Qualifiers

```

source
1..3465
/organism="unidentified"
/db_xref="taxon:32644"
/clone="PHS2BIOS1"
272..1477
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB77089.1"
/db_xref="GI:7240888"
/translation="MNVFNPQFRAQFPALQDAGVYLDNATLALKEPAVEATQOFYS
LSAGNVHRSQFAEORLTARVEAAREKVAOQLNAPDQKTIWTRGTESINVAOCYA
RPLQPGDEIIVVAHNAHNPMLVAOOGAKVYKPLNAPRLPDVLLPLITPR
SRILALGOMSNVTGGCPDLARATTFPAHSAGMYVNVGAOGAVHPADVOOLDIDFAF
SGHKLVGPTGIVLXKSELLEMSWLGKGVNHYVSPGCTQSAPMKLEAGTQNV
AGVIGLSAALWELADYDINOESMSRLATLADALAKRPFSPFCQDSSLLAFDPA
GVHSDMVTLLAEYGIALRAGHCAQPLLAELGVTLRASFAPYNTKSDVALVNAV
DRALLELYVD"

BASE COUNT      804 a      963 c      934 g      764 t
ORIGIN
Query Match      100.0%; Score 1206; DB 6; Length 3465;
Best Local Similarity 100.0%; Pred. No. 1.1e-301;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAACGTTTAAATCCCGGCACTTTCGGCCAGTTTCCCGGCACTACAGATGCGGGC 60
    |||||||
DB 272 ATGAACGTTTAAATCCCGGCACTTTCGGCCAGTTTCCCGGCACTACAGATGCGGGC 331

OY 61 GTCTATCTCGACAGCGCGGCGACGCGCTTAAACCTGAAGCCGTGTTGAAGCCACCA 120
    |||||||
DB 332 GTCTATCTCGACAGCGCGGCGACGCGCTTAAACCTGAAGCCGTGTTGAAGCCACCA 391

OY 121 CAGTTTACAGTCTGAGCGCGGGAAGCTGCATGCGACCCAGTTTCCCGAAGCCCAAGC 180
    |||||||
DB 392 CAGTTTACAGTCTGAGCGCGGGAAGCTGCATGCGACCCAGTTTCCCGAAGCCCAAGC 451

OY 181 CTGACCGCGGCTTATGAAGTGCACGAGAGAAGTGGCGGCAATTAATGAATGACCGGAT 240
    |||||||
DB 452 CTGACCGCGGCTTATGAAGTGCACGAGAGAAGTGGCGGCAATTAATGAATGACCGGAT 511

OY 241 GATAAACTATGCTGAGACGCGGCGACCACTGAATCCATCAACATGCTGGCAATGCG 300
    |||||||
DB 512 GATAAACTATGCTGAGACGCGGCGACCACTGAATCCATCAACATGCTGGCAATGCG 571

OY 301 TATCGCGGCTCGGCTCTGCAACCGCGGCGATGAGATTATGTCACCGTGGCAACACCAC 360
    |||||||
DB 572 TATCGCGGCTCGGCTCTGCAACCGCGGCGATGAGATTATGTCACCGTGGCAACACCAC 631

OY 361 GCAACGCTGCGGCTGATGCTGCGCCCAACAACTGAGGCAAAAGTGGTGAATG 420
    |||||||
DB 632 GCAACGCTGCGGCTGATGCTGCGCCCAACAACTGAGGCAAAAGTGGTGAATG 691

OY 421 CCGCTTAAATGCGACGCACTCGCGATGTCGATTTGTTCCAGAACTGATTAATCCCGCT 480
    |||||||
DB 692 CCGCTTAAATGCGACGCACTCGCGATGTCGATTTGTTCCAGAACTGATTAATCCCGCT 751

OY 481 AGTGGGATTGCGGCTGGGTCAGATGCGAACCTTACTGGCGGTTGGCCGATCTGGCG 540
    |||||||
DB 752 AGTGGGATTGCGGCTGGGTCAGATGCGAACCTTACTGGCGGTTGGCCGATCTGGCG 811

OY 541 CGACGATTAATGCTGCTCATTCAGCCGCGATGCTGATGCTGATGCTGCTCAGGGG 600
    |||||||
DB 812 CGACGATTAATGCTGCTCATTCAGCCGCGATGCTGATGCTGATGCTGCTCAGGGG 871

OY 601 CGAGTGCATTTCCCGCGGATGTTCCAGCACTGATGATGATTTCTATGCTTTTCAAGT 660
    |||||||
DB 872 CGAGTGCATTTCCCGCGGATGTTCCAGCACTGATGATGATTTCTATGCTTTTCAAGT 931

OY 661 CACAACGTAATGCGCGGCGAGAGTATGCGGCTGATGCTGATGCTGATGCTGCTGAG 720
    |||||||
DB 932 CACAACGTAATGCGCGGCGAGAGTATGCGGCTGATGCTGATGCTGATGCTGCTGAG 991

OY 721 GCGATGTCGCCCTGGCTGGCGGCGCAAAATGTTTCAAGAGTGAATTTTGACGCGCTTC 780

```

```

DB 992 GCGATGTCGCCCTGGCTGGCGGCGGCGCAAAATGTTTACAGAGTGAATTTTACGCGCTTC 1051
OY 781 ACGACTCAATCTGCGCGCGTGAAGAACTGGAAGCTGGAAGCCCAATGCTGCTGCATATA 840
DB 1052 ACGACTCAATCTGCGCGCGTGAAGAACTGGAAGCTGGAAGCCCAATGCTGCTGCATATA 1111
OY 841 GGAATTAAGCCCGCGGCTGGAATGCTGCGAGATTACATATTCAAACCGCCGCAAACTGG 900
DB 1112 GGATTAAGCCCGCGGCTGGAATGCTGCGAGATTACATATTCAAACCGCCGCAAACTGG 1171
OY 901 AGCCGTAGCTTAGCAACGCTGGCGGAGATGCGCTGGCGGAAGCTCCCGCTTTCCTTCA 960
DB 1172 AGCCGTAGCTTAGCAACGCTGGCGGAGATGCGCTGGCGGAAGCTCCCGCTTTCCTTCA 1231
OY 961 TTCCGCTGCAAGATTCACCGCTGCTGCGCTTTCATTTTCCTGCGCTTCATAGCGAT 1020
DB 1232 TTCCGCTGCAAGATTCACCGCTGCTGCGCTTTCATTTTCCTGCGCTTCATAGCGAT 1291
OY 1021 ATGCTGACGCTGCTGGCGGAGATGCGATTTGCCCTGGCGGCGGCGAGCATGCGCTCAG 1080
DB 1292 ATGCTGACGCTGCTGGCGGAGATGCGATTTGCCCTGGCGGCGGCGAGCATGCGCTCAG 1351
OY 1081 CCGCTACTGCGAGAAATTAAGCGCTAACCGCACACTGCGCGCTCTTTTTCGCGCATATAAT 1140
DB 1352 CCGCTACTGCGAGAAATTAAGCGCTAACCGCACACTGCGCGCTCTTTTTCGCGCATATAAT 1411
OY 1141 ACAAGAGTATGATGATGCGCTGCTGTAATGCCCTTACCGCGCGCTGGAATTAATGCTG 1200
DB 1412 ACAAGAGTATGATGATGCGCTGCTGTAATGCCCTTACCGCGCGCTGGAATTAATGCTG 1471
OY 1201 GATTAA 1206
DB 1472 GATTAA 1477

RESULT 4
AX000474      3794 bp      DNA      linear      PAT 10-MAR-2000
LOCUS      AX000474
DEFINITION      Sequence 5 from Patent WO9905285.
ACCESSION      AX000474
VERSION      AX000474.1 GI:7240883
KEYWORDS
SOURCE      unidentified.
ORGANISM      unidentified.
REFERENCE      1 (bases 1 to 3794)
AUTHORS      Schroeder,H. and Hauer,B.
TITLE      METHOD FOR PRODUCING BIOTIN
JOURNAL      Patent: WO 9905285-A 5 04-FEB-1999;
              SCHROEDER HARTWIG (DE); BASIP AG (DE)
              Location/Qualifiers:
                source
                  1..3794
                    /organism="unidentified"
                    /db_xref="taxon:32644"
                    /clone="PHS2BIOS1"
                    601..1806
                      /note="unnamed protein product"
                      /codon_start=1
                      /protein_id="CAB77089.1"
                      /db_xref="GI:7240888"
                      /translation="MNVFNPQFRAQFPALQDAGVYLDNATLALKEPAVEATQOFYS
                      LSAGNVHRSQFAEORLTARVEAAREKVAOQLNAPDQKTIWTRGTESINVAOCYA
                      RPLQPGDEIIVVAHNAHNPMLVAOOGAKVYKPLNAPRLPDVLLPLITPR
                      SRILALGOMSNVTGGCPDLARATTFPAHSAGMYVNVGAOGAVHPADVOOLDIDFAF
                      SGHKLVGPTGIVLXKSELLEMSWLGKGVNHYVSPGCTQSAPMKLEAGTQNV
                      AGVIGLSAALWELADYDINOESMSRLATLADALAKRPFSPFCQDSSLLAFDPA
                      GVHSDMVTLLAEYGIALRAGHCAQPLLAELGVTLRASFAPYNTKSDVALVNAV
                      DRALLELYVD"

BASE COUNT      910 a      1034 c      1020 g      830 t
ORIGIN
Query Match      100.0%; Score 1206; DB 6; Length 3794;

```

Best Local similarity 100.0%; Pred. No. 1.le-301;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAACGTTTTTAATCCGCGAGTTTGGCGCCGATTTCCGCGCACTACAGATGCGGCG 60
Db 601 ATGAACGTTTTTAATCCGCGAGTTTGGCGCCGATTTCCGCGCACTACAGATGCGGCG 660
OY 61 GTCTATCTGACAGCGCGCGACCGCGCTTAACCTGAAGCGGTGGAACCCACCA 120
Db 661 GTCTATCTGACAGCGCGCGACCGCGCTTAACCTGAAGCGGTGGAACCCACCA 720
OY 121 CAGTTTACAGTCTGACGCGCGGAATGCTCATGCGAGCTTTGCGGACCCCAAGCG 180
Db 721 CAGTTTACAGTCTGACGCGCGGAATGCTCATGCGAGCTTTGCGGACCCCAAGCG 780
OY 181 CTGACCGCGCGTTATGAAGCTGCAAGAGAAAGTGGCCATTTCTGATGACCGCGAT 240
Db 781 CTGACCGCGCGTTATGAAGCTGCAAGAGAAAGTGGCCATTTCTGATGACCGCGAT 840
OY 241 GATTAACACTATGCTGACGCGCGGACCACTGATCATCAATGATGCTGCGACATGCG 300
Db 841 GATTAACACTATGCTGACGCGCGGACCACTGATCATCAATGATGCTGCGACATGCG 900
OY 301 TATGCGCGCTCGCGTCTGCAACCGGCGGATGATTAATTTGTGAGCGGTGCGAGACAC 360
Db 901 TATGCGCGCTCGCGTCTGCAACCGGCGGATGATTAATTTGTGAGCGGTGCGAGACAC 960
OY 361 GCGAACCTGCTGCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 961 GCGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
OY 421 CCGCTTAATGCGCGACGAGTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 1021 CCGCTTAATGCGCGACGAGTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
OY 481 AGTCGGATTCGCGCTGCGTGGGTGAGATGTCGAACGTTACTGCGCGTTGCCGAGTCTG 540
Db 1081 AGTCGGATTCGCGCTGCGTGGGTGAGATGTCGAACGTTACTGCGCGTTGCCGAGTCTG 1140
OY 541 CGAGCGATTAACCTTTGCTCATACCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 1141 CGAGCGATTAACCTTTGCTCATACCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
OY 601 GCAAGTCAATTTCCCGCGGATGCTGCACTGATGATTTGATTTCTGCTTTTCAGCT 660
Db 1201 GCAAGTCAATTTCCCGCGGATGCTGCACTGATGATTTGATTTCTGCTTTTCAGCT 1260
OY 661 CACAACCTGATGCGCGCAGAGTATGCGCTGCTGATGTAATCAGAACTGCTGAG 720
Db 1261 CACAACCTGATGCGCGCAGAGTATGCGCTGCTGATGTAATCAGAACTGCTGAG 1320
OY 721 GCGATGCTGCGCTGCTGCGCGCGGCAAAATGTTTACAGAACTGATTTTACGCGCTTC 780
Db 1321 GCGATGCTGCGCTGCTGCGCGCGGCAAAATGTTTACAGAACTGATTTTACGCGCTTC 1380
OY 781 ACAGACTCAATCTGCGCGCTGCAAACTGGAAGCTGGAAGCCCAATGCTGCTGCTGCTA 840
Db 1381 ACAGACTCAATCTGCGCGCTGCAAACTGGAAGCTGGAAGCCCAATGCTGCTGCTGCTA 1440
OY 841 GGAATTAAGCGCGCGCTGGAATGCTGCGAGATTAACGATATCAACGAGCGCAAACTG 900
Db 1441 GGAATTAAGCGCGCGCTGGAATGCTGCGAGATTAACGATATCAACGAGCGCAAACTG 1500
OY 901 AGCCGTAAGCTTGAACAGCTGCGGGAAGATGCGCTGCGCAAGCTGCGCGCTTTCCTTA 960
Db 1501 AGCCGTAAGCTTGAACAGCTGCGGGAAGATGCGCTGCGCAAGCTGCGCGCTTTCCTTA 1560
OY 961 TTCGCGTGCAGAGATTCAGCGCGCTGCGCTTGTGATTTGCTGCGCTTATCATCATAGCAT 1020
Db 1561 TTCGCGTGCAGAGATTCAGCGCGCTGCGCTTGTGATTTGCTGCGCTTATCATCATAGCAT 1620
OY 1021 ATGTCAGCCTGCTGCGGAGATAGGATTAATGCTGCGCGCGGAGCATGCTGCTGAG 1080

```

Db 1621 ATGTCAGCCTGCTGCGGAGTACGATTAATGCTGCGCGCGCGGCGGCGGCGGCGGCG 1680
OY 1081 CCGTACTGCGGAGATTAATGCGGCTACGACACTGCGCGCTCTTTTGGCGCATATAT 1140
Db 1681 CCGTACTGCGGAGATTAATGCGGCTACGACACTGCGCGCTCTTTTGGCGCATATAT 1740
OY 1141 ACMAAGAGTATGATGCGCTGCTGTAATGCGGCTTGAACCGCGCGCTGGAATTAATGCTG 1200
Db 1741 ACMAAGAGTATGATGCGCTGCTGTAATGCGGCTTGAACCGCGCGCTGGAATTAATGCTG 1800
OY 1201 GATTA 1206
Db 1801 GATTA 1806

RESULT 5
AX018970
LOCUS AX018970 3794 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 11 from Patent WO9942591.
ACCESSION AX018970
VERSION AX018970.1 GI:10043065
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia coli.
REFERENCE Schroeder H.
METHOD Method for producing bioclin
PATENT Patent: WO 9942591-A 11 26-AUG-1999;
JOURNAL SCHROEDER HARTWIG (DE); BASF AG (DE)
LOCATION/Qualifiers:
1. 3794
/organism="Escherichia coli"
/db_xref="taxon:562"
601. 1806
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07719.1"
/db_xref="GI:10043066"
/translation="MNVFNPAPQAFAPALDAGVYLDSAAATALKPEAVVEATQOFYS
LSAGNVHRSQFAEAOFLPAPAEAREKVAOLNLPDDTITVTRCTESIMVAVCYA
RPLQPDDEITVSVAENHNLVPMYLMQGAQVKVLPNAOLRPDVLPELTPR
SRILAGMSNVGSGPDLARATFPAISAGVAVVVDGQAVHPADVOQDIDIFYA
SGHLKPTGIGVLYKSELLEAMSPHIGGKRNHVEFDDFTTOSAPWKLEACPYN
AGVIGLSALEMLADYDINQAESWRSLATLEBDLARRPGRFRSCODSSILAFDA
GVHSDVNTLEAEYGLALRAQOHCAPLAEGLVGTLLRASFPAPYNTKSDVDALVNAV
DRALELVD"

BASE COUNT 910 a 1034 c 1020 g 830 t
ORIGIN

Query Match 100.0%; Score 1206; DB 6; Length 3794;
Best Local similarity 100.0%; Pred. No. 1.le-301;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAACGTTTTTAATCCGCGAGTTTGGCGCCGATTTCCGCGCACTACAGATGCGGCG 60
Db 601 ATGAACGTTTTTAATCCGCGAGTTTGGCGCCGATTTCCGCGCACTACAGATGCGGCG 660
OY 61 GTCTATCTGACAGCGCGCGACCGCGCTTAACCTGAAGCGGTGGAACCCACCA 120
Db 661 GTCTATCTGACAGCGCGCGACCGCGCTTAACCTGAAGCGGTGGAACCCACCA 720
OY 121 CAGTTTACAGTCTGACGCGCGGAATGCTCATGCGAGCTTTGCGGACCCCAAGCG 180
Db 721 CAGTTTACAGTCTGACGCGCGGAATGCTCATGCGAGCTTTGCGGACCCCAAGCG 780
OY 181 CTGACCGCGCGTTATGAAGCTGCAAGAGAAAGTGGCCATTTCTGATGACCGCGAT 240
Db 781 CTGACCGCGCGTTATGAAGCTGCAAGAGAAAGTGGCCATTTCTGATGACCGCGAT 840
OY 241 GATTAACACTATGCTGACGCGCGGACCACTGATCATCAATGATGCTGCGACATGCG 300

```



```

Db 841 GATAAACTATGCTGTGAGACGGCGGACATCAATCCATCAACATGTTGGCCATTC 900
Oy 301 TATGCGGCTCGCGCTGTCGAACCGGGGAGATGATTTATTTGACCGCTGGAGAACACAC 360
Db 901 TATGCGGCTCGCGCTGTCGAACCGGGGAGATGATTTATTTGACCGCTGGAGAACACAC 960
Oy 361 GCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 961 GCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Oy 421 CCGCTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 1021 CCGCTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Oy 481 AGTCGATTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 1081 AGTCGATTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Oy 541 CGAGCGATTTACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 1141 CGAGCGATTTACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Oy 601 GCAGTGCATTTCCCGCGGAGTGTTCAGCAACTGATGATTTGCTGCTGCTGCTGCTGCTGCT 660
Db 1201 GCAGTGCATTTCCCGCGGAGTGTTCAGCAACTGATGATTTGCTGCTGCTGCTGCTGCTG 1260
Oy 661 CACAACGTATGAGCGCGGAGAGTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 1261 CACAACGTATGAGCGCGGAGAGTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Oy 721 GCGATGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 1321 GCGATGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Oy 781 ACAGCTCAATCTGCGCGCTGGAACACTGGAAGCTGGAACGCGCAATGCTGCTGCTGCTGCT 840
Db 1381 ACAGCTCAATCTGCGCGCTGGAACACTGGAAGCTGGAACGCGCAATGCTGCTGCTGCTGCT 1440
Oy 841 GGATTAAGCGCGGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAAT 900
Db 1441 GGATTAAGCGCGGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAAT 1500
Oy 901 AGCGGTAGCTTACGACGCGGCGGAGAGATGCGCTGGCGAAGCTCCCGGCTTGGTTCA 960
Db 1501 AGCGGTAGCTTACGACGCGGCGGAGAGATGCGCTGGCGAAGCTCCCGGCTTGGTTCA 1560
Oy 961 TTCCGCTGCGGAGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 1561 TTCCGCTGCGGAGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Oy 1021 ATGCTGACGCTGCTGCGGAGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1621 ATGCTGACGCTGCTGCGGAGATTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
Oy 1081 CCGCTACTGCGAGATTAAGCGGTAACCGGCACTGGCGGCTTTTGGCCCAATTAAT 1140
Db 1681 CCGCTACTGCGAGATTAAGCGGTAACCGGCACTGGCGGCTTTTGGCCCAATTAAT 1740
Oy 1141 ACAAGAGTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 1741 ACAAGAGTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
Oy 1201 GATTAA 1206
Db 1801 GATTAA 1806

```

```

RESULT 6
AX018972 4975 bp DNA linear PAT 07-SEP-2000
LOCUS AX018972
DEFINITION Sequence 13 from Patent WO942591.
ACCESSION AX018972

```

```

VERSION AX018972.1 GI:10043067
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
REFERENCE Schroeder, H.
AUTHORS Method for producing biotin
TITLE Patent: WO 9942591-A 13 26-AUG-1999;
JOURNAL SCHROEDER, HARTWIG (DE); BASF AG (DE)
FEATURES
source location/Qualifiers
1..4975
/organism="Escherichia coli"
/db_xref="taxon:562"
530..1684
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07720.1"
/db_xref="GI:10043068"
/transl_table="MANIFESTESVSEGHDPKIDAOISDAVDAILEDDPKARVACET
YKTMGVLVGGEITTSAYVDIEITRNTRREIGYVSHSDGFANSCAVLSALCKOSP
INQVGRADPLBOGADGILMGVATNETNDVLPAPITTAHLVROAEVRKNGTLP
LRPDAKSOVTFQYDDSKYIGDAVYVSTHSEIDLOKSLQEMELIKPIPLAEML
SATKFINPTGRFVIGIPMGDCGLTRKTIYDPTGGMARHGGAFFSGKDPKVBREA
YARIVAKNIVAAGLADKCEIOVSTAIGAEPTSIWETFGTEKVSBDLVVREF
DLRPGLIQLMLDLPIYKETAAYGHFHEHPWEXTDKAOLLRAAGLK"
1782..2987
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07721.1"
/db_xref="GI:10043069"
/transl_table="MANIFESTESVSEGHDPKIDAOISDAVDAILEDDPKARVACET
YKTMGVLVGGEITTSAYVDIEITRNTRREIGYVSHSDGFANSCAVLSALCKOSP
INQVGRADPLBOGADGILMGVATNETNDVLPAPITTAHLVROAEVRKNGTLP
LRPDAKSOVTFQYDDSKYIGDAVYVSTHSEIDLOKSLQEMELIKPIPLAEML
SATKFINPTGRFVIGIPMGDCGLTRKTIYDPTGGMARHGGAFFSGKDPKVBREA
YARIVAKNIVAAGLADKCEIOVSTAIGAEPTSIWETFGTEKVSBDLVVREF
DLRPGLIQLMLDLPIYKETAAYGHFHEHPWEXTDKAOLLRAAGLK"
1782..2987
BASE COUNT 1198 a 1350 c 1329 g 1098 t
ORIGIN
Query Match 100.0%; Score 1206; DB 6; Length 4975;
Best Local Similarity 100.0%; Pred. No. 1..301; Indels 0; Caps 0;
Matches 1206; Conservative 0; Mismatches 0;
Oy 1 ATGACCTTTTAAATCCCGGCAATTCGCGCCCACTTTCGCCACATCAAGATCCGGC 60
Db 1782 ATGACCTTTTAAATCCCGGCAATTCGCGCCCACTTTCGCCACATCAAGATCCGGC 1841
Oy 61 GTCATATCGACAGCGCGGAGCGGCTTAAACCTGAACCGGCTGTAAGCCACCAA 120
Db 1842 GTCATATCGACAGCGCGGAGCGGCTTAAACCTGAACCGGCTGTAAGCCACCAA 1901
Oy 121 CAGTTTACAGTCTGACGCGCGGAACGTCATCGCAGCCAGTTTGGCGAAGCCCAAGC 180
Db 1902 CAGTTTACAGTCTGACGCGCGGAACGTCATCGCAGCCAGTTTGGCGAAGCCCAAGC 1961
Oy 181 CTGACCGCGGCTTATGAACTGCGACGAGAAAGTGGCGCAATTACGATGCAACCGGAT 240
Db 1962 CTGACCGCGGCTTATGAACTGCGACGAGAAAGTGGCGCAATTACGATGCAACCGGAT 2021
Oy 241 GATTAACATATCGTCTGAGCGCGGAGCACTGAATCCATCAACATGCTGCGCAATTC 300
Db 2022 GATTAACATATCGTCTGAGCGCGGAGCACTGAATCCATCAACATGCTGCGCAATTC 2081
Oy 301 TATGCGGCTCGCGCTGTGCAACCGGCGAGATGATTTGCTGACGCTGCGAGAACACAC 360
Db 2082 TATGCGGCTCGCGCTGTGCAACCGGCGAGATGATTTGCTGACGCTGCGAGAACACAC 2141

```

```

OY 361 GCCAACCTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCAAAGTGGTAATTC 420
    |||||||
DB 2142 GCCAACTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCAAAGTGGTAATTC 2201
OY 421 CCGCTTAATGCGCAGCAGCAGTCCCGGATGTCGATTTGTTGCCAGAACTGATCTCCCGT 480
    |||||||
DB 2202 CCGCTTAATGCGCAGCAGCAGTCCCGGATGTCGATTTGTTGCCAGAACTGATCTCCCGT 2261
OY 481 ACTCGGATTCGCGCTTGGGTGATGTCGAACTTACTGGCGTTGCCGATCTGGCG 540
    |||||||
DB 2262 ACTCGGATTCGCGCTTGGGTGATGTCGAACTTACTGGCGTTGCCGATCTGGCG 2321
OY 541 CGAGGATTAACCTTTGCTATTACAGCCGGGATGTCGATGCTGATGCTGCTGAGG 600
    |||||||
DB 2322 CGAGGATTAACCTTTGCTATTACAGCCGGGATGTCGATGCTGATGCTGCTGAGG 2381
OY 601 CGAGTGCATTTCCCGCGGATGTTCCAGCAACTGATGATTTGATTTCTTTCAGGT 660
    |||||||
DB 2382 CGAGTGCATTTCCCGCGGATGTTCCAGCAACTGATGATTTGATTTCTTTCAGGT 2441
OY 661 CACAACCTGATGTCGCCGAGCAGATATCGGCTGCTGATGTCGATGTCGATGTCGAG 720
    |||||||
DB 2442 CACAACCTGATGTCGCCGAGCAGATATCGGCTGCTGATGTCGATGTCGATGTCGAG 2501
OY 721 CGGATGTCGCCGCTGCTGCGCGGCGGCAAAATGTTTACAGAACTGATTTGACGGCTTC 780
    |||||||
DB 2502 CGGATGTCGCCGCTGCTGCGCGGCGGCAAAATGTTTACAGAACTGATTTGACGGCTTC 2561
OY 781 ACAGCTCAATCTGCGCGCTGGAACCTGGAACCTGGAACCTGGAACCTGCTGCTGCTGATA 840
    |||||||
DB 2562 ACAGCTCAATCTGCGCGCTGGAACCTGGAACCTGGAACCTGGAACCTGCTGCTGCTGATA 2621
OY 841 GGAATTAAGCGCGCGCTGGAATGCTGCGAGATTAAGATTAACCAACCGCCGCAAACTGCG 900
    |||||||
DB 2622 GGAATTAAGCGCGCGCTGGAATGCTGCGAGATTAAGATTAACCAACCGCCGCAAACTGCG 2681
OY 901 AGCGGATTAACCTGGAACCTGGAACCTGGAACCTGGAACCTGCTGCTGCTGCTGATA 960
    |||||||
DB 2682 AGCGGATTAACCTGGAACCTGGAACCTGGAACCTGGAACCTGCTGCTGCTGCTGATA 2741
OY 961 TTCCGCTGCGCAGATTCAGACCTGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 1020
    |||||||
DB 2742 TTCCGCTGCGCAGATTCAGACCTGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 2801
OY 1021 ATGCTGACCTGCTGCGGAGATGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 1080
    |||||||
DB 2802 ATGCTGACCTGCTGCGGAGATGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 2861
OY 1081 CCGCTACTGCGCAATTAAGCGGATGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 1140
    |||||||
DB 2862 CCGCTACTGCGCAATTAAGCGGATGCTGCGGAGATGCTGCGGCGGCGGCGGCTTCTGCTCA 2921
OY 1141 ACAAGAGTATGTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
    |||||||
DB 2922 ACAAGAGTATGTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2981
OY 1201 GATTAA 1206
    |||||||
DB 2982 GATTAA 2987
    |||||||

RESULT 7
AE000364 12058 bp DNA linear BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 254 of 400 of the complete
DEFINITION genome.
ACCESSION AE000364.1 GI:2367162
VERSION AE000364.1 GI:2367162
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 12058)
          Blatterner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
          Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
          Gregor, J., Davis, N.W., Kirpatrick, H.A., Goeden, M.A., Rose, D.J.,
          Mau, B., and Shao, Y.
          The complete genome sequence of Escherichia coli K-12
          Science 277 (5331), 1453-1474 (1997)
JOURNAL 97428503
MEDLINE 9278503
PUBMED 9278503
REFERENCE
AUTHORS 2 (bases 1 to 12058)
          Blatterner, F.R.
          Direct Submission
          Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
          3 (bases 1 to 12058)
          Blatterner, F.R.
          Direct Submission
          Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
          4 (bases 1 to 12058)
          Plunkett, G. III.
          Direct Submission
          Submitted (13-OCT-1998) Laboratory of Genetics, University of
          Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
          On Sep 9, 1997 this sequence version replaced gi:1789165.
          This sequence was determined by the E. coli Genome Project at the
          University of Wisconsin-Madison (Frederick R. Blattner, director).
          Supported by NIH grants HG00301 and HG01428 (from the Human Genome
          Project and NCHGR). The entire sequence was independently
          determined from E. coli K12 strain MG1655. Predicted open reading
          frames were determined using Genemark software, kindly supplied by
          Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
          30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
          have been correlated with genetic loci are being annotated with CG
          Site Nos., unique ID nos. for the genes in the E. coli Genetic
          Stock Center (CGSC) database at Yale University, kindly supplied by
          Mary Berlyn. A public version of the database is accessible
          (http://cgsc.biology.yale.edu). Annotation of the genome is an
          ongoing task whose goal is to make the genome sequence more useful
          by correlating it with other data. Comments to the authors are
          appreciated. Updated information will be available at the E. coli
          Genome Project's World Wide Web site
          (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
          its annotations are periodically updated; this is Version M54. No
          sequence changes are periodically updated; updated gene identifications
          and products: all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.

FEATURES
source Location/Qualifiers
    1..12058
        /organism="Escherichia coli K12"
        /strain="K12"
        /sub_strain="MG1655"
        /db_xref="taxon:83333"
        45..67
            /note="central position to predicted promoter: -231..5"
            /bound_molecule="Aarg predicted site"
            252..282
                /note="factor sigma70: predicted +1 start at 2932242"
                304..1620
                    /gene="fucP"
                    /note="b2801"
                    304..11620
                        /gene="fucP"
                        /function="transport; Transport of small molecules:

```

Carbohydrates, organic acids, alcohols"
/note="0438: 100 pct identical to FUCP_ECOLI SW: P11551;
CG site no. 10875"
/codon_start=1
/transl_table=11
/product="fucose permease"
/protein_id="AAC75843.1"
/db_xref="GI:1789166"
/translation="MGNTSIQTOSYRAVDKAGSRSTYIPALLCSLEFLMAVANL
NDLLPQFOAFTLTFQAGLISAFYEFYFIPPIAGILMKLSYKAGITIGLEFLA
LGAALFPAEIMNTYTLFVGLFIAGAGCLETAPANFVTLGPSSSHRPLNLT
FNSFGIIVVVGOSLISNVPHOSODVLDKMSPOLSAVKSLVLSVOTPMIIVAI
VLLVALLMLTFEPALOSDNDKAGSSASLSLARIRHMAVLAQFCVAGQOTA
CWSYLRIVAEITPCMTAGFANNTLTGTWCCFTGCTGTWLISRAPKVLAAALI
AAALCLISAFAGHGLIALILCSAFMSIOTYFISLGIKNLGDTKYGSSTVMTII
GGIYTPVWGFSDAAGNIPTELIPALCFAVIFAFRRSQTATN"
1653. .3428
/gene="fucI"
/note="b2802"
1653. .3428
/gene="fucI"
/EC_number="5.3.1.-"
/function="enzyme: Degradation of small molecules: Carbon
compounds"
/note="0591: 100 pct identical to FUCI_ECOLI SW: P11552;
CG site no. 10878"
/codon_start=1
/transl_table=11
/product="L-fucose isomerase"
/protein_id="AAC75844.1"
/db_xref="GI:1789167"
/translation="MKKISLPKIGIRPVIDGRMGVRESLEQTNMKAATAALTEK
LRHACGAVECVISDTCIAGMAEAACEKFSQNGLTITVPCWCYSEITIDMPT
RRAKMGNGTERPAVYLAALAHAOSKIPAFSLYGHVDADDTSPADVEELK
RRARAGLAVASMKKSYSLGCGSMGIAQIVDHFESMLCMKVOAVMTELSRRID
OKIYDAELEMLAMADKNRYGDENNKOYONASRAVLRSLTLMACIRDMQ
NSKIADIGVEESIGYNAIAGGOGORHWTDOYNGDTAAILNSGDFMNGREPVY
ATENDSLNCAVHLMGHQTLGTNOVPADVTYSPAEIRVETGKLDGLAHEGIIHIN
SSAALDSCSKODESGNPTMKPHEISQOEADACLAETMCPALHEYRGGYSRF
LTEGGVPTMTRVNIIKGLGVLQIAEGMSVELPKVDHDLINKRTNSTVPTWFAFL
TGKGPTDYSVANMNGANHGVLITIGVAGDEFITLASHMLRIPVCMHNVETKYRPSA
MAAHMDIGDQDRAQNYGPLYKR"
3507. .4955
/gene="fucK"
/note="b2803"
3507. .4955
/gene="fucK"
/EC_number="2.7.1.51"
/function="enzyme: Degradation of small molecules: Carbon
compounds"
/note="0482: 100 pct identical to FUCI_ECOLI SW: P11553;
CG site no. 19878"
/codon_start=1
/transl_table=11
/product="L-fuculokinase"
/protein_id="AAC75845.1"
/db_xref="GI:1789168"
/translation="MLSGYIAGAIMKQEVILVDCATNVAIAVNRCKIVARASTP
NASDIAMENNTIHOHSLDALIORFADCCROINSELECHIRGIAVTFGVDGLVNO
GNLLYPLISMKCPRTAAYMDNIERLISABRLOISGVGAFSEMTLKKVLMKRNHOL
LEBAHAHLFISLILNRLTGEFTDTITMAGTSMIDIOQRDSPOLQATGIRRLFP
RLVEAGEQIGTIONSAAAMLGLPVGIPIVLSAGHDTFALGAGAEONEPVLSCGTET
LWVRSQVDTLSLSOYAGTCELDISOAGLVNPMQMLASGVLEWKRLLPMTACTPM
LIEEARLIRAGADVKNOCDDLSCONAGMCGVLTNTRGHIFYRAALEGJTAOLORILQ
MLEIGHFASSELLVGGSRNTLMNOICANMMDIPVKVLDAETTVAGAAALFGMYGV
GFENSPBEARAQIHQYRYFYPTPEPEFIEEV"
4957. .5379
/gene="fucU"
/note="b2804"
4957. .5379
/gene="fucU"
/function="phenotype: Degradation of small molecules:
Carbon compounds"

/note="0140: 100 pct identical to FUCU_ECOLI SW: P11555
but35 C-terminal aa are dissimilar to reported 28 aa"
/codon_start=1
/transl_table=11
/product="protein of fucose operon"
/protein_id="AAC75846.1"
/db_xref="GI:1789169"
/translation="MLKTIISPLISPELLKVAEMGHGDEIIFSDAHFPAHSGPROVIR
ADGLVSDLOLAIPLFFLDSTAPPVMAAAYEGDTLQEVERIRNALSLQAPCPDI
IRINFAVEYERAKAFAYITGERKAYGNILLKRGVTP"
5437. .6168
/gene="fucR"
/note="b2805"
5437. .6168
/gene="fucR"
/function="regulator: Degradation of small molecules:
Carbon compounds"
/note="0243: 100 pct identical to FUCR_ECOLI SW: P11554;
CG site no. 10884"
/codon_start=1
/transl_table=11
/product="positive regulator of the fuc operon"
/protein_id="AAC75847.1"
/db_xref="GI:1789170"
/translation="MKAAKROAIVDLLNHTSLTEALSEOLKYSKETIRDLNELQ
TCKTILRNHRAKYIHRONDSDGPFHRLKSHYAKRADARALAWIEGMYIAIDAS
STCWYLAHQLPDINTOVFNHPIHEKREKRIOLISSGTLERKGYVPSLSISO
LKSLEIDLFHIFESCEGIDSGLAMDNSNAIADKSMILKRAAOSLLLIDSKFNRSGEA
RIGHIDEVTHIISDEROYATSLVTA"
complement(6212. .7312)
/gene="ygdE"
/note="b2806"
complement(6212. .7312)
/gene="ygdE"
/function="orf: unknown"
/note="f366: 100 pct identical to YGDE_ECOLI SW: P32066;
alternate name orf: of X73413 and U01030"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75848.1"
/db_xref="GI:1789171"
/translation="MNKVVLLCRPRGPEKCAAEIDKAGOREIFGARYKENAGVYI
ECYOPDDCKLIRELPFSFLIFAROMFYVGLLOHLPREDRTITPIYGLQGVENKGE
Query Match 100.0%; Score 1206; DB 1; Length 12058;
Best Local Similarity 100.0%; Pred. No. 1,1e-301;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 1 ATGACGTTTAAATCCCGCGGAGTTTCGGCCCACTTCCGCACTACAGATCCGGCC 60
DB 9406 ATGACGTTTAAATCCCGCGGAGTTTCGGCCCACTTCCGCACTACAGATCCGGCC 9465
QY 61 GTCTATCTGCACAGCGCGCGGACCGGCTTAAACGTGAAGCGGTGTAAGCCACCA 120
DB 9466 GTCTATCTGCACAGCGCGCGGACCGGCTTAAACGTGAAGCGGTGTAAGCCACCA 9525
QY 121 CAGTTTACAGTGTGAGCGCGCGGAAACGTGCATCGCAGCCAGTTTCCGAAGCCCAAC 180
DB 9526 CAGTTTACAGTGTGAGCGCGCGGAAACGTGCATCGCAGCCAGTTTCCGAAGCCCAAC 9585
QY 181 CTGACCGCGCGGTATGACGTGCACAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
DB 9586 CTGACCGCGCGGTATGACGTGCACAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 9645
QY 241 GATAAACATATGCTGTGAGCGCGCGGACAGCACTGAATCATCAACATGATGGCACAATGC 300
DB 9646 GATAAACATATGCTGTGAGCGCGCGGACAGCACTGAATCATCAACATGATGGCACAATGC 9705
QY 301 TATGCGCGTCCGCGGTGCAACCGCGCGGATAGATTATTGTCAAGCGTGGCAGAACACAC 360
DB 9706 TATGCGCGTCCGCGGTGCAACCGCGCGGATAGATTATTGTCAAGCGTGGCAGAACAC 9765

OY 361 GCCAACCTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCCAAGTGTGAATTG 420
 DB 9766 GCCAACCTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCCAAGTGTGAATTG 9825
 OY 421 CCCCTTAATGCGAGCGACGATGCGGATGTGATTTTGGCCAGACTGATTTCCCGT 480
 DB 9826 CCCCTTAATGCGAGCGACGATGCGGATGTGATTTTGGCCAGACTGATTTCCCGT 9885
 OY 481 AGTCGGATTTCTGGCGTGGTGCATGATGTGAACCTACTGCGGCTGGCCGATCTGGCG 540
 DB 9886 AGTCGGATTTCTGGCGTGGTGCATGATGTGAACCTACTGCGGCTGGCCGATCTGGCG 9945
 OY 541 CGAGCATTAACCTTTGCTCTATTACCGCGGATGCTGATGCTGATGCTGCTACGCG 600
 DB 9946 CGAGCATTAACCTTTGCTCTATTACCGCGGATGCTGATGCTGATGCTGCTACGCG 10005
 OY 601 GCAGTCAATTTCCCGCGGATGTTGAGCAACTGGAATTTGATTTTCAATCTTTTACGT 660
 DB 10006 GCAGTCAATTTCCCGCGGATGTTGAGCAACTGGAATTTGATTTTCAATCTTTTACGT 10065
 OY 661 CACAACCTGATGCGCGGACAGGATGCGGCTGATGCTGATGCTGATGCTGATGCTGAG 720
 DB 10066 CACAACCTGATGCGCGGACAGGATGCGGCTGATGCTGATGCTGATGCTGATGCTGAG 10125
 OY 721 GCGATGCGCGCTGCTGCGGCGCGCAAAATGTTTACAGACTGATTTTACGCGCTT 780
 DB 10126 GCGATGCGCGCTGCTGCGGCGCGCAAAATGTTTACAGACTGATTTTACGCGCTT 10185
 OY 781 ACCACTCAATCTGCGCGCTGGAACCTGGAACCTGGAACCGCAATGCTGCTGCTGCTA 840
 DB 10186 ACCACTCAATCTGCGCGCTGGAACCTGGAACCTGGAACCGCAATGCTGCTGCTGCTA 10245
 OY 841 GGAATTAAGCGCGCGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTG 900
 DB 10246 GGAATTAAGCGCGCGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTGGAATGCTG 10305
 OY 901 ACCCTGAGTTTGAACAACGCTGCGGGAAGATGCGCTGCGGCAACGCTGCGGCTTTCCTCA 960
 DB 10306 ACCCTGAGTTTGAACAACGCTGCGGGAAGATGCGCTGCGGCAACGCTGCGGCTTTCCTCA 10365
 OY 961 TTCCTGCGCAGAGATTCACAGCTGCTGCGCTTGTGATTTTCTGCGCTTCAATACGAT 1020
 DB 10366 TTCCTGCGCAGAGATTCACAGCTGCTGCGCTTGTGATTTTCTGCGCTTCAATACGAT 10425
 OY 1021 ATGCTGACCTGCTGCGGAGATGCTGCTGCTGCTGCGGCGCGGAGCATTTGCGCTCAG 1080
 DB 10426 ATGCTGACCTGCTGCGGAGATGCTGCTGCTGCTGCGGCGCGGAGCATTTGCGCTCAG 10485
 OY 1081 CCGCTACTGCGCAATTAAGCGCTAACCGGACACTGCGGCTCTTTTGGCCCATTAAT 1140
 DB 10486 CCGCTACTGCGCAATTAAGCGCTAACCGGACACTGCGGCTCTTTTGGCCCATTAAT 10545
 OY 1141 ACAAGAGTATGATGATGCGCTGCTGATGCTGCTGATGCGCGGCTGCAATTAATGCTG 1200
 DB 10546 ACAAGAGTATGATGATGCGCTGCTGATGCTGCTGATGCGCGGCTGCAATTAATGCTG 10605
 OY 1201 GATTAA 1206
 DB 10606 GATTAA 10611

RESULT 8

ECUC29581 71128 bp DNA linear BCT 14-JAN-1997
 DEFINITION Escherichia coli K-12 genome: approximately 63 to 64 minutes.
 ACCESSION U29581
 VERSION U29581.1 GI:1255723

KEYWORDS Escherichia coli.
 SOURCE Escherichia coli.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 71128)

AUTHORS Reizer, J., Reizer, A., Merrick, M.J., Plunkett, G. III, Rose, D.J. and Sailer, M.H. Jr.
 TITLE Novel phosphotransferase-encoding genes revealed by analysis of the Escherichia coli genome: a chimeric gene encoding an Enzyme I homologue that possesses a putative sensory transduction domain
 JOURNAL 97128/75
 MEDLINE 8973315
 PUMED 2 (bases 1 to 71128)
 REFERENCE Plunkett, G.
 AUTHORS Direct Submission
 TITLE Submitted (20-JUN-1995) Guy Plunkett, University of Wisconsin, Laboratory of Genetics, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT On Apr 5, 1996 this sequence version replaced GI:882681.
 This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. This entry should be considered provisional.
 Reference [1] describes a correction and update to the original entry's sequence and annotations.
 Location/Qualifiers
 1..71128
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 /map="approximately 63 to 64 minutes"
 /note="This sequence comprises the following lambda clones: DD628(EC18-104), DD632(EC30K458-4), DD635(EC18-251), DD639(EC30MM20), and DD641(EC17MR8); M13Janus vector was used for subcloning"
 complement(1..940)
 /note="ORF_1313p"
 /codon_start=1
 /transl_table=11
 /protein_id="AAB40437.1"
 /db_xref="GI:882682"
 /translation="MSQSTTTPVVEEMOYIPVAGHDSMLNLSGAHPFTNTIYLIK DNGHTVGEIPGKERIKRTLEDALPLVVGKLGKRLNLTIVRNTPADRDGKQLO TFDLRTTHVYVIEAAMLDLGOHLGVNVASLDGDOORSEVHLGLTFGNKRAT PLPYOSDPDSCDMYRLRIEEMTPDAVRLAAYEKGFDFLKGGLVLEGEAE SIVALAORPQARITLDPGAGMSLEAIKIGKYLAVLADYADPCAGEGFGSREVMA EERRATGLPTATNMATDPRONGHLSLDSVDIIRRI"
 complement(961..2101)
 /note="ORF_1446"
 /codon_start=1
 /transl_table=11
 /protein_id="AAB40438.1"
 /db_xref="GI:882683"
 /translation="MATQSSPYITDMKVIPVAGHDSMLNLSGAHNAFTNIVLTD NAGHTGIGEPGQDYIYOTLVDAIPVHLOEYARLKLKYYOQVHKNOADPTFEKCA WFEELRVNVALEALLLTLGLKALNPVCELGLGKQKEALTVLGLYIDRRTKD LPYEVNTPGNHMYQLRHCKANSESAVVALEASDRYGFNFKLKGGLPGEODEIDT VVALKRPEDNARTIVDPNGALNLDLDEAI SLCKLNDVLTVAEDPCAGEGFGSREVMA FRATGCLPVATNMIATNMEEMGHAVLANVDI PLADPFWMTLSGAVRYVADLDDMGLT WCHSNHNHFDISLAFTHYVAGAPGNPTPIIDTHMIWQEDCLTQNPLEIKKXIAVP DAPGLGVELDMEOVQKAHAYRRLPGCARNDGCPHOYLIPGNTFDRKRVRFGRH"
 complement(2303..3655)
 /note="ORF_1450"
 /codon_start=1
 /transl_table=11
 /protein_id="AAB40439.1"
 /db_xref="GI:882684"
 /translation="MSSLSQAASVSEKRTNAYIVVLMLEIYTSFNGDRAATLSIAGS EAKDIGLDPGQGVVFSFSAVAYIGQIPGMLDLDRFGSKRVYFMSJFIMWFTLLO GVDIEFSGFGLI VALFETLRLFLVGLAEPSPGNLSIVAMPAPQORGPVLSFNSAOY FATVIFAPIMGLTLHEVMSHVEFFPMGAGIYISPIMLKIVFEPHONPVNKELEYI AAGCALINDDONTKRVKPSVYKGOIQLDLSRMIGYIGYOXYINATYFEIIFMP VYLVQNRCHSILKAGFVAVPACGFIGVLCGCTSDMLMARTGSLNTRAKPIYMKH LLSMVFNCNVNVEVEMIIIGFMALEFEGIGALDAMVADTAPEISGLSGCLNMF GNISCIIVPIAIGYIVGTGTFNGALIVGVHALLAVLSLVLVGDIDRIELKPVAGO

CDS

CDS
complement(4090..4539)
/note="corresponds to hypothetical protein from E. carotovora PIR Accession Number S45107; ORF_f150"
/codon_start=1
/transl_table=11
/protein_id="AAB40440.1"
/db_xref="GI:882685"
/translation="MAEIGFVGTGYNGLVAEAEALITAGGKATVEDEPESLDM LPQDYVLTSTGGDLPDSIVPLFGIKDGLGPNLRGVIALDSSVYVNCN GKKFALLQEOSAORVEMLLIDASENEPEPESNPWEQNTLLS"
complement(4557..5339)
/note="corresponds to hypothetical protein from E. carotovora PIR Accession Number S45108; 2 ORFs in E. coli; ORF_f260"
/codon_start=1
/transl_table=11
/protein_id="AAB40441.1"
/db_xref="GI:882686"
/translation="MLEILYODEMLVAVKNKPSGMLVHRSLDREKVVVQVTRDQIG QHVFTHRLDRPTSGVLLGLSEAGRLLAQOEQIQKRYHAIVRGMLEAVDY PLVEELDKIADKAREDKGPOPAVTRGLATYEMVATGRVPTTRGLVLELPKGR KHLRRLHLAHLRPIITGDSKKGDLQNRSGAEHFGLORLMLHASQLSLTHPTGCELT IHAGLDPTWQALSTQSGMRGLLEPNERVESAPSGDGEISS"
complement(5339..5668)
/note="corresponds to hypothetical protein from E. carotovora PIR Accession Number S45108; 2 ORFs in E. coli; ORF_f109"
/codon_start=1
/transl_table=11
/protein_id="AAB40442.1"
/db_xref="GI:882687"
/translation="MTTHDRVRLQLOALREHOMRDEPOPHQFNSTOPFFMDT MELEWLOWLIPRMHDLDNKQPLPGAFVAVYEMALATDHPQALILAELEKIDA LFMDDAS"
complement(6290..6835)
/note="GTG start"
/codon_start=1
/transl_table=11
/protein_id="AAB40443.1"
/db_xref="GI:882688"
/translation="MDLITAOALKDFATRYCDAMHEHKSMPLESELYGVSPCIIST TEDAVYHQPOPTGEONVNAVERAFDIVIOTPHRTTTPQFADKMAQRPDIITLQ TWSEDFRVOENLIGHLVOKRLKPLTLEATLEELVISVCMLSGEVCAETLGT RKRTHLASNALELNQKPL"
6903..7751
/note="ORF_0282"
/codon_start=1
/transl_table=11
/protein_id="AAB40444.1"
/db_xref="GI:882689"
/translation="MSSVYANHQAAGLTLGKSHYRDYDASLLQGVPRSLNDPGL KADNLPFGHDITWLYELSWLNKGLPOVAVGVGLDYTSVNLIESKSFLYNSFNO TRNNMDEVYROTLEKDLSTCAQKISVALYRLDELGQRPAGNCIDODDITINPA EPTTDLNATCEKCYEBETLVSLKSNCLITHQDPMGSLQIQYGRQIDREKLLRY LVSFRHNEFHQCVKIFENDLRFQPEKLSVARYTRKGLDINPMSNDVPST RLVRO"
7863..9227
/note="TTC start; alternate gene name ygdH; ORF_0455"
/codon_start=1
/transl_table=11
/protein_id="AAB40445.1"
/db_xref="GI:882690"
/translation="MITHISPLGSMDSLEVDMLKRTASSDLYOLFNRCSLAVLNS GSLDNKSKLLSFENEDINVLRRREGVKLELINPEEAFVDRIIRALQANFAVLR DILFYVGOIHNTVRPNLNDNSVHTLVESFLLRNARALHVEGARPMVYVCMGHSIN ENFYLVARRYGNOGLRELNICTGCGSGAGAEAPKKAAGAAQORXKDSFTEMTPS IIAEPNPVLNELLITMPDIKLEAFVRIASHIILFPGVGAEELVILGLILMIPA NKDOVPLILITGKESADYFRVLDENVYHLLGNARHRIIIDDAAEVAROKKSNP LKNEKRDQDAYSFNMSMRILADLPMPFSPENNAANLKLVDQVDEVAADLRRAF SGIVAGNVKEVGIKRAIEEGPYKINDCKEIMRMDLLQFVAQHRMKLPGSAYICY EICT"
misc_feature
8982..11136

gene
/note="similar to GenBank Accession Number U01233 (BCU01233) 1..2135"
9784..11073
/gene="sdac"
9784..11073
/gene="sdac"
/note="CG Site no. 33329; ORF_0429"
/codon_start=1
/transl_table=11
/product="putative serine transporter"
/protein_id="AAB40446.1"
/db_xref="GI:882691"
/translation="METTOTSTIASKDSANRKTPTMMLGICRAIACVFLPIN AGVGMIPILIMAILAFPMPTFFAHRGLTRFVLSGKPGEDITEVEHFGICAGLIT LLYFAIYPLILVSVATINTVESFMSHOLGTPPRALLSLILVGMNTIVRFGEM IYKAMSLIVFEVGVMLLALYILPOMNGAALLETSLDPAATGNLMTMLAIIPM VFSFNSPLISSFAVAKREYGDMAOKSKSLFAHIMVLTVPFVSVCVSLTPA DLAAREONISILSYLAHFNAPVIAAMPAITAIITSPGLHYGAREGFGVYIK SLRQKRSIEINKRITZLPHLVTTWIVATLNPSTLGMIEPLGPIIMILFLPMY AIQKVPAMKRYSGHISNVFVWGLIAISATYLSFS"
10731..13515
/note="similar to GenBank Accession Number L07763 (BCOSDAB) 1..2785"
11131..12498
/gene="sdab"
11131..12498
/gene="sdab"
/EC_number="4.2.1.13"
/note="CG Site no. 33324; ORF_0455"
/codon_start=1
/transl_table=11
/product="L-serine dehydratase 2 (L-serine deaminase 2)"
/protein_id="AAB40447.1"
/db_xref="GI:882692"
/translation="MISVFDFRIGIGIPSSHTVGPMAKQFTDOLIARNLKDVT R VYDVYGSISLTGKHDIALIMGLAGLPPHVDSDISPSIIOVNTGRMLANGO HEYEPVDDCMNPHADNLSLHENGKRITRLADKVVYSSOTYSISGGFVYDEHNGO DSAIPVEVPYSSAADLKHCOETGLSLGMMKNKLAHSKELEBOLHANWEYWRG GIERGISTEGVLSGLKLRVRAAALRLMLVSDKTTDPAVAVDWIMNPLAVNEENA AGGRVYATPTNGACGIIIFAVLAYDFFIREVYANSLARYLVAASIGSLYKMASSG AEVGCQGEVGNSSMAAGLAEILLGASPPQVCIAGIAGIENHNLGLTCDVAVOVYPC IERNADIAKAAVNAARMLRRTSEPVCLDKVLETWYETGKQMNNAKVRTSGGLAMK IYACD"
12520..13365
CDS
100.0%; Score 1206; DB 1; Length 71126;
Best Local Similarity 100.0%; Pfed. No. 1.2e-301;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Query Match
1 ATGACGTTTAAATCCCGCGGAGTTTGGCGCCAGTTTCCGCACTACAGATCGCGG 60
24892 ATGACGTTTAAATCCCGCGGAGTTTGGCGCCAGTTTCCGCACTACAGATCGCGG 24951
DB
61 GTCTATCTGACAGCGCGCGGAGCGGCTTAAACGTGAAGCGGCTTGAAGCACC 120
24952 GTCTATCTGACAGCGCGCGGAGCGGCTTAAACGTGAAGCGGCTTGAAGCACC 25011
DB
121 CAGTTTACAGTGTGACCGCGGGAACGTTCATCGCAGCCAGTTTGGCGAAGCCCAACG 180
25012 CAGTTTACAGTGTGACCGCGGGAACGTTCATCGCAGCCAGTTTGGCGAAGCCCAACG 25071
DB
181 CTGACCGCGGCTTATGAGCTGCACAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
25072 CTGACCGCGGCTTATGAGCTGCACAGAAAGTGGCGCAATTACTGAATGCACCGGAT 25131
DB
241 GATTAACACTATCGCTGAGCGCGGCGGACAGCTAGTAATCCATCAACATGAGGCGCAATTC 300
25132 GATTAACACTATCGCTGAGCGCGGCGGACAGCTAGTAATCCATCAACATGAGGCGCAATTC 300
25133 GATTAACACTATCGCTGAGCGCGGCGGACAGCTAGTAATCCATCAACATGAGGCGCAATTC 300
DB
301 TATGCGCGTGGCGGCTGCAACCGCGGCGATGAGATTATGTCAGCGGTGGCAGAACACAC 360
25192 TATGCGCGTGGCGGCTGCAACCGCGGCGATGAGATTATGTCAGCGGTGGCAGAACACAC 360

Oy 361 GCCAACCTCGTCCCTGCTGATGTCGCCCAACAACTGAGCCAAAGTGTGAATTTG 420
 |||||||
 Db 25352 GCCAACCTCGTCCCTGCTGATGTCGCCCAACAACTGAGCCAAAGTGTGAATTTG 25311
 Oy 421 CCGCTTAATCGCAGCAGCAGCTGCGGATGTGATTTGTTGCCAGACTGATTCTCCCGT 480
 |||||||
 Db 25312 CCGCTTAATCGCAGCAGCAGCTGCGGATGTGATTTGTTGCCAGACTGATTCTCCCGT 25371
 Oy 481 AGTCGGATTGTGGGTTGGTGCAGATGTGACAGCTTACTGGCGGTCCCGGATCTGGCG 540
 |||||||
 Db 25372 AGTCGGATTGTGGGTTGGTGCAGATGTGACAGCTTACTGGCGGTCCCGGATCTGGCG 25431
 Oy 541 CGAGCGATTACCTTTGCTCATTTACGCCGAGTGTGATGTTGATGTCACAGCG 600
 |||||||
 Db 25432 CGAGCGATTACCTTTGCTCATTTACGCCGAGTGTGATGTTGATGTCACAGCG 25491
 Oy 601 GCACTGATTTCCCGCGGATGTTTACAGCACTGATATTTGATTTCTATGCTTTTACAGT 660
 |||||||
 Db 25492 GCACTGATTTCCCGCGGATGTTTACAGCACTGATATTTGATTTCTATGCTTTTACAGT 25551
 Oy 661 CACAACTGATGCGCGGAGATGCGGCGTGTATGTAATGAACTGCTGGAG 720
 |||||||
 Db 25552 CACAACTGATGCGCGGAGATGCGGCGTGTATGTAATGAACTGCTGGAG 25611
 Oy 721 GCGATGTCCGCTGCTGCTGCGCGCGCAAAATGTTTACAGAGTATTGACGCTTC 780
 |||||||
 Db 25612 GCGATGTCCGCTGCTGCTGCGCGCGCAAAATGTTTACAGAGTATTGACGCTTC 25671
 Oy 781 ACGACTCAATCTGCGCGCTGGAACAGTGAAGCTGCAAGCCCAATGTCGCTGTCTATA 840
 |||||||
 Db 25672 ACGACTCAATCTGCGCGCTGGAACAGTGAAGCTGCAAGCCCAATGTCGCTGTCTATA 25731
 Oy 841 GGATTAAGCGCGCGCGGAGATGCTGGCAGATATACATACACAGCGCGAAGCTGG 900
 |||||||
 Db 25732 GGATTAAGCGCGCGGAGATGCTGGCAGATATACATACACAGCGCGAAGCTGG 25791
 Oy 901 AGCGGTAGCTTACGACGCTGCGCGAGAGATGCGCTGGCGAAGCTCCCGGCTTCTGTTCA 960
 |||||||
 Db 25792 AGCGGTAGCTTACGACGCTGCGCGAGAGATGCGCTGGCGAAGCTCCCGGCTTCTGTTCA 25851
 Oy 961 TTCCGCTGCGCAGATTTCCAGCTCTGCTGCTTATTTTCTGCTGCTTATCATACAGCAT 1020
 |||||||
 Db 25852 TTCCGCTGCGCAGATTTCCAGCTCTGCTGCTTATTTTCTGCTGCTTATCATACAGCAT 25911
 Oy 1021 ATGTGACGCTGCTGCGGAGATGCTGCTGCTGCTGCGGCGCGGAGCATTTGGCTCATACAGCAT 1080
 |||||||
 Db 25912 ATGTGACGCTGCTGCGGAGATGCTGCTGCTGCTGCGGCGCGGAGCATTTGGCTCATACAGCAT 25971
 Oy 1081 CCGCTACTGCGAGATTTAGCGGTAAACCGCAGACTGCGCGCTCTTTTGGCCCATATTAAT 1140
 |||||||
 Db 25972 CCGCTACTGCGAGATTTAGCGGTAAACCGCAGACTGCGCGCTCTTTTGGCCCATATTAAT 26031
 Oy 1141 ACAAGAGTATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 |||||||
 Db 26032 ACAAGAGTATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 26091
 Oy 1201 GATTAA 1206
 |||||||
 Db 26092 GATTAA 26097
 |||||||
 RESULT 9
 AX000470
 LOCUS AX000470 1216 bp DNA linear PAT 10-MAR-2000
 DEFINITION Sequence 1 from Patent WO9905285.
 ACCESSION AX000470
 VERSION AX000470.1 GI:7240879
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 1216)

AUTHORS Schroeder, H. and Hauer, B.
 TITLE METHOD FOR PRODUCING BIOTIN
 JOURNAL Patent: WO 9905285-A 1 04-FEB-1999;
 SCHROEDER, HARTWIG (DE); BASF AG (DE)
 FEATURES
 source
 1. '1216
 /organism="Escherichia coli"
 /strain="W3110"
 /db_xref="taxon:552"
 12. '>1216
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CA877085.1"
 /db_xref="GI:7240880"
 /translation="MVNPADQRAQFPALDQAGVLDLSNATLAKPEAVYEATQRTS
 LSAQVNRQFAEQR/LTAREYAEAREKVAQOLLAAPDQKTIWRTGTTESINVAQCYA
 RPLQPDDELIVSAEHNANLTPMLVVAQDQKVKVLP/LNAQRLPDVQDLP/LITPR
 SRLALGQSNVSGCPDLARATTFPAHSAGMVMVVDGAGAVHF PADVQDLDIFVAF
 SGHKLXGPTGIGVLYGKSELLEAMSPLGGKVMVHVSFTDGTQSAQPKLEAGTQNV
 AGVIGLSALEMLADYDINQAESMSLQATLADALAKRPGSPFCQDSLLAFQPA
 GVHSDVTLLEIETGIALNRGQHCAPLAELEQVGTLRASFAPYNTKSDVDALVNAV
 DRALLELVQ"
 BASE COUNT 258 a 318 c 363 g 277 t
 ORIGIN
 Query Match 99.9%: Score 1205; DB 6; Length 1216;
 Best Local Similarity 100.0%: Pred. No. 1.8e-301;
 Matches 1205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ATGACGTTTTTAATCCCGCAGATTTCGCCCGCAGTTCCCGCAGTACAGATGCGCG 60
 |||||||
 Db 12 ATGACGTTTTTAATCCCGCAGATTTCGCCCGCAGTTCCCGCAGTACAGATGCGCG 71
 Oy 61 GTCTATCTGACAGCGCGCGGACCGCGCTTAACCTGAAGCGTGTGAAGCCACCAA 120
 |||||||
 Db 72 GTCTATCTGACAGCGCGCGGACCGCGCTTAACCTGAAGCGTGTGAAGCCACCAA 131
 Oy 121 CAGTTTACAGTCTGACCGCGCGGAAAGTCCATGCGAGCCGTTTGGCGAAGCCCAACGC 180
 |||||||
 Db 132 CAGTTTACAGTCTGACCGCGCGGAAAGTCCATGCGAGCCGTTTGGCGAAGCCCAACGC 191
 Oy 131 CTGACCGCGCGCTTATGAAGCTGACAGAGAGAAGTGGCGCAATTAATGAATGACACCGAT 240
 |||||||
 Db 192 CTGACCGCGCGCTTATGAAGCTGACAGAGAGAAGTGGCGCAATTAATGAATGACACCGAT 251
 Oy 241 GATAAACTATGCTGTGAGCGCGCGGACCACTGAATTCATCAACATGCTGCAACATGC 300
 |||||||
 Db 252 GATAAACTATGCTGTGAGCGCGCGGACCACTGAATTCATCAACATGCTGCAACATGC 311
 Oy 301 TATCGCGGTCCGCTGTCGCAACCGCGGATGAGATTAATTCACAGTGGCGCAACACAC 360
 |||||||
 Db 312 TATCGCGGTCCGCTGTCGCAACCGCGGATGAGATTAATTCACAGTGGCGCAACACAC 371
 Oy 361 GCCAACCTGTCGCCCTGCTGATGTCGCCCAACAACTGAGCAGCAAAAGTGTGAATTTG 420
 |||||||
 Db 372 GCCAACCTGTCGCCCTGCTGATGTCGCCCAACAACTGAGCAGCAAAAGTGTGAATTTG 431
 Oy 421 CCGCTTAATGCGCAGCAGCTGCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCGT 480
 |||||||
 Db 432 CCGCTTAATGCGCAGCAGCTGCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCGT 491
 Oy 481 AGTCGGATTCTGGGCTTGGTCAATGTGCAAGCTTACTGGCGGTGGCCCGGATCTGGCG 540
 |||||||
 Db 492 AGTCGGATTCTGGGCTTGGTCAATGTGCAAGCTTACTGGCGGTGGCCCGGATCTGGCG 551
 Oy 541 CGAGGATTACCTTTGCTCATTTACAGCGCGAGATGCGGATGTTGATGCTGCTACAGCG 600
 |||||||
 Db 552 CGAGGATTACCTTTGCTCATTTACAGCGCGAGATGCGGATGTTGATGCTGCTACAGCG 611
 Oy 601 GCACTGCAATTTCCCGCGGATGTTACAGCACTGATATGATTTCTATGCTTTTACAGT 660
 |||||||

Db 612 GCAGTCATTTCCCGCGGATGTCAGCACTGATATTGATTTCTATGCTTTTCAGCT 671

Qy 661 CACAACTGTATGCCCCGACAGTATCGCGCTGTATGTAATCAGAACTGCTGAG 720

Db 672 CACAACTGTATGCCCCGACAGTATCGCGCTGTATGTAATCAGAACTGCTGAG 731

Qy 721 GCGATGTCGCGCTGCTGGCGCGGCAAAATGCTTCACAGAGTGTATTTACCGCTTC 780

Db 732 GCGATGTCGCGCTGCTGGCGCGGCAAAATGCTTCACAGAGTGTATTTACCGCTTC 791

Qy 781 ACAGCTCATATGCGCGCGGCAAACTGGAAGCTGGAACCCCAATGCTGCTGTGCATA 840

Db 792 ACAGCTCATATGCGCGCGGCAAACTGGAAGCTGGAACCCCAATGCTGCTGTGCATA 851

Qy 841 GCATTAAAGCGCGCGCTGGAATGCTGCGAGATTACAGATATCAGACGCCCAAAAGCTGCG 900

Db 852 GCATTAAAGCGCGCGCTGGAATGCTGCGAGATTACAGATATCAGACGCCCAAAAGCTGCG 911

Qy 901 ACCCGTACGTTAGCAACGCTGGCGGAAGATGGCGTGGCAACGTCGCCGCTTGTCTCA 960

Db 912 ACCCGTACGTTAGCAACGCTGGCGGAAGATGGCGTGGCAACGTCGCCGCTTGTCTCA 971

Qy 961 TTCGCTGCGCAGATTTCCAGCCTGCTGCGCTTGTATTTTGTGCGCTTCATCATAGCAT 1020

Db 972 TTCGCTGCGCAGATTTCCAGCCTGCTGCGCTTGTATTTTGTGCGCTTCATCATAGCAT 1031

Qy 1021 ATGTGACGCTGCTGCGCGAGTACGCTATTCCTGCGCGCGCGCAGCATTTGCGCTCAG 1080

Db 1032 ATGTGACGCTGCTGCGCGAGTACGCTATTCCTGCGCGCGCGCAGCATTTGCGCTCAG 1091

Qy 1081 CCGCTACTGCGCAGATTTAGCGCTTAACCGCGCAGTGGCGCGCTCTTTTGGCGCATATAT 1140

Db 1092 CCGCTACTGCGCAGATTTAGCGCTTAACCGCGCAGTGGCGCGCTCTTTTGGCGCATATAT 1151

Qy 1141 ACAAGAGTATGTGATGCGCTGCTGTAATGCCCTGACCCGCCCTGCAATTTATTTGTTG 1200

Db 1152 ACAAGAGTATGTGATGCGCTGCTGTAATGCCCTGATGATGCCGTTGACCCGCCCTGCAATTTATTTGTTG 1211

Qy 1201 GATTA 1205

Db 1212 GATTA 1216

RESULT 10

AP002563 26658 bp DNA linear BCT 07-MAR-2001

LOCUS Escherichia coli O157:H7 DNA, complete genome, section 14/20.

DEFINITION AP002563 BA000007

ACCESSION AP002563.1 GI:13363121

VERSION

KEYWORDS

SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)

ORGANISM Escherichia coli O157:H7

REFERENCE Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae; Escherichia.

AUTHORS 1 (sites)

1 (sites)

2 (sites)

Ohnishi, M., Murata, T., Nakayama, K., Kihara, S., Hattori, M., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and Shinagawa, H.

Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)

20198780

TITLE

JOURNAL

REFERENCE

AUTHORS

Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an

Escherichia coli K-12 strain MG1655

Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

MEDLINE

20557356

3 (sites)

Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and Shinagawa, H.

Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak

Gene 258 (1-2), 127-139 (2000)

20564182

4 (sites)

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsuda, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kihara, S., Shibata, T., Hattori, M., and Shinagawa, H.

Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12

DNA Res. 8 (1), 11-22 (2001)

21156231

5 (bases 1 to 26658)

Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.

Direct Submission

Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)

genome project.

COMMENT

FEATURES

Source

1..26658

/organism="Escherichia coli O157:H7"

/strain="O157:H7"

/sub-strain="RIMD 0509952"

/db_xref="taxon:83334"

/complement(244..693)

/gene="ECS3650"

/complement(244..693)

/gene="ECS3650"

/note="similar to B2790.ECOLI g111789154 percent identity 100 in 149 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not-experimental

/transl_table=1

/product="hypothetical protein"

/protein_id="BAB37073.1"

/db_xref="GI:13363122"

/translation="MAETGIFVGTMGNSLVAEAEATLTAGCHKATVEDEPDLSDM LPYQDKVLVTSTYGGDLPSDIPVLFQGIKSLGFPNLRVGVIALGDSYVNFEN GGRQFDALLOEQAORVGEMLLIDASENPETESNPWEOMCTLLS"

/complement(711..1493)

/gene="ECS3651"

/complement(711..1493)

/gene="ECS3651"

/note="similar to YOCB.ECOLI g111789155 percent identity 99 in 260 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not-experimental

/transl_table=1

/product="hypothetical protein"

/protein_id="BAB37074.1"

/db_xref="GI:13363123"

/translation="MLELIVODEMLVANKPSGLVHRSWLDDEKVVVQTVRODQIG OHVFTARLDPRPTSGVLLKGLSSAGRLAQQEFOHIOKRYHAIYVGHLMEEAVLDY PLVEELQIDKDFAREDKGPOAVTHRGATVEMPAATGRTPTRGVLELPKPCR KQRLRLALHRLPITDCKHGDRLRQNSAHRFGQRLMLHLSLSLTHPPFGCEPLT IHAGLDITWQALSQFGWGLDPENRVERVSADSCGDGERS"

/complement(1493..1822)

/gene="ECS3652"

/complement(1493..1822)

/gene="ECS3652"

/gene="ECS3652"

gene

CDS


```

/note="similar to B2792_ECOLI g111789156 percent identity
98 in 109 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37075.1"
/db_xref="GI:13363124"
/transl_table=11
/translation="MTTHDRVRLQALQALREHOMRWDEPOHOFNSTOPEFMDT
MEPLEHLOWYLIRMDLNNQPLPGAFVARYEMALATDHPORALLIAELEKIDA
LFADAS"
/gene="ECs3653"
/complement(1865..2233)
/gene="csfB"
/complement(1865..2233)
/gene="csfB"
/product="CsfB RNA"
/note="Inhibitor of CsfA"
/complement(2444..2989)
/gene="ECs3653"
/complement(2444..2989)
/gene="ECs3653"
/note="similar to SYD_ECOLI g111789157 percent identity 98
in 181 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="Syd protein"
/protein_id="BAB37076.1"
/db_xref="GI:13363125"
/transl_table=11
/translation="MDLLTQALKDFARYCDAMHEEKKSMPLSEELYGVSPCIIST
TDAVYWKQPOPFGGEONVAVERAFDIVIOPTIHTFTTQFAGDMAQFCDITLLQ
TMSDDPRRYOENLIGLITOKRLKPLPTLFATLEBELEVISVCNLSGEVCKETLGT
RKRTHLAPNLAFLNLQKPL"
/gene="ECs3654"
/3057..3905
/gene="ECs3654"
/note="similar to YOCB_ECOLI g111789158 percent identity
98 in 282 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:13363126"
/transl_table=11
/translation="MSSTYANQALAGLTLCKSTIDYRDYTSALLQGYPRSLNDPLCL
KADLPFGQTDITLYELSWLNKGLPQVAVGVLEDTYSVNLIESKSFLLYNSFNO
TRFNMDREVROTLELDLSTCAGEVSVALRLDELEQPIGHNGTCIDODDITIDY
EFTTDYLENATSEKVEETLVSHLKSNCNLTHOPMGSIOIYOYRGROIDREKLELY
LVSPRHNEHNEOCVERIFMDLRFQCPKELSYARTRRGGLDINWRSNDSPVST
TRYRO"
/gene="ECs3655"
/4017..5381
/gene="ECs3655"
/4017..5381
/gene="ECs3655"
/note="similar to YGDH_ECOLI g111789159 percent identity
99 in 454 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37078.1"
/db_xref="GI:13363127"
/transl_table=11
/translation="MTHISPLGSMQMLSOLEVMDLKRKTASSDLYOLFNCSLAVLS
GSLDNKSLLSRENFIDIVLRERGVKLELINPEEAVVDGRIIALDANLFAVIR
DILFYGOIHNTVAFPALNDNSVHITNLFSILRNARALHVGAPMVMQWCHGHSIN
ENEYLYARVNOGLRELNICGCGGAPKKAAGAAVHAOORXYDSRFIGMTEPS
IIAEPNPPLVNLIIIMPDIEKLEAFVRIAHGIIIPGCGTAELLYLIGLIMNA
NKDOYPLILITGKRESADYFRVDEPVRHTLGEAKRHHYIIIDDAEVAVRQKKKMP
LVKENRDTGDATSFMSRIADPLOWPEPSSHNMANLKLIPQDPVEVLAADLRRAF
SGIAGVNWKEVGIIRALIEFGPYKINGDKIMRMDLLOFVADHRMKLPGSAVIICY
EICH"
/gene="ECs3656"
/5938..7227

```

```

Query Match 96.8% Score 1167.6; DB1; Length 26658;
Best Local Similarity 98.0%; Pred. No. 1.2e-91;
Matches 1182; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

CDS
/gene="ECs3656"
/5938..7227
/note="similar to SDAC_ECOLI g111789160 percent identity
100 in 429 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="probable serine transporter"
/protein_id="BAB37079.1"
/db_xref="GI:13363128"
/transl_table=11
/translation="METTOTSTIASKDSRSRKRTDMMGLGYTAIGAVLFLPIN
ACGVGAPLPIILAIAPMTFARHGLTRFVLSGKPGEDITEVEEHGIGKGLIT
LYFAIYPLILVSVATNTVESFMSHOGPPRAIISLIIIGMMTYVFGOM
IVKAMSILVPPFVGLMLATLYLIPOMNALETSLSDPASATGNLMTMLAIPIVM
VSPNSPIISSFAVAKRETEGMAOKSKILAFHIMVLYMFPVSCVILTPA
DLAAKQONISILYLAHFNAPVIAHMAPIIAITTSFLGHVIGARSGFNGVIK
SLRGKRSIEINKLRITLFLMLVTWIIAYATLPSIETLGGP1IAM1LPLMPY
AIOKVPAMRYSGHISNVFVVGIIAISIFYSIFS"
/gene="ECs3657"
/7285..8652
/gene="ECs3657"
/note="similar to SDAA_ECOLI g111789161 percent identity
77 in 455 aa, similar to SDAB_ECOLI g111789161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"

CDS
/gene="ECs3657"
/7285..8652

```

Db 21646 GCGGTGCAATTTCCCGGCGATGTCAGCAACTGATCGATTCTCTATGCTTTCTCTGGT 21705
 Oy 661 CACAACACTGATGCCCCGACAGGATTCGCGCTGCTGTAATGTAATACAACTGCTGGAG 720
 Db 21706 CACAACACTGATGCCCCGACAGGATTCGCGCTGCTGTAATGTAATACAACTGCTGGAG 21765
 Oy 721 GCGATGTCGCGCTGCGGCGGCAAAATGTTACGAAGTGAAGTTTGGAGCGCTTC 780
 Db 21766 GCGATGTCGCGCTGCGGCGGCAAAATGTTACGAAGTGAAGTTTGGAGCGCTTC 21825
 Oy 781 ACGACTCAATCTGCGCGCTGGAAGCTGGAACGCCCAAAATGTCGCTGCTCATTA 840
 Db 21826 ACGACTCAATCTGCGCGCTGGAAGCTGGAACGCCCAAAATGTCGCTGCTCATTA 21885
 Oy 841 GGATTAACGCGCGCGCTGGAATGCTGCGAGATTACGATATACACAGCGCCGAAGCTTCG 900
 Db 21886 GGATTAACGCGCGCGCTGGAATGCTGCGAGATTACGATATACACAGCGCCGAAGCTTCG 21945
 Oy 901 ACGCGTAGCTTAGCAACGCTGGGGAAGATGCGGTGGGAACGTCGCGCTTGTCTCA 960
 Db 21946 ACGCGTAGCTTAGCAACGCTGGGGAAGATGCGGTGGGAACGTCGCGCTTGTCTCA 22005
 Oy 961 TTCGCTGCGACAGATTCCAGCCTGCTGCGCTTGTGATTTTGTGCGCTTCATCATAGCGAT 1020
 Db 22006 TTCGCTGCGACAGATTCCAGCCTGCTGCGCTTGTGATTTTGTGCGCTTCATCATAGCGAT 22065
 Oy 1021 ATGCTGACGCTCTGCGGAGATACGCTATTCGCTGCGGCGCGGCGGACAGATTGCGCTCAG 1080
 Db 22066 ATGCTGACGCTCTGCGGAGATACGCTATTCGCTGCGGCGCGGCGGACAGATTGCGCTCAG 22125
 Oy 1081 CCGCTACTGCGAGATTAGCGCTGACCGGACACTGCGCGCTCTTTGGCCCATATAAT 1140
 Db 22126 CCGCTACTGCGAGATTAGCGCTGACCGGACACTGCGCGCTCTTTGGCCCATATAAT 22185
 Oy 1141 ACAAAAGTAGATGTGATGCGCTGCTGTAATGCGCTTGACCGCGCTGGAATTATTGCTG 1200
 Db 22186 ACAAAAGTAGATGTGATGCGCTGCTGTAATGCGCTTGACCGCGCTGGAATTATTGCTG 22245
 Oy 1201 GATTAA 1206
 Db 22246 GATTAA 22251

RESULT 11
 AE005509
 LOCUS 6488 bp DNA linear BCT 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 128
 ACCESSION AE005509 AE005174
 VERSION AE005509.1 GI:12517290
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
 1 (bases 1 to 6488)
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamoudis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.
 Genome sequence of enterohemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 FEATURES
 source

Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 Location/Qualifiers
 1..6488
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155664"
 /note="enterohemorrhagic"
 complement(206..436)
 /gene="24126"
 complement(206..436)
 /gene="24126"
 /function="orf: Unknown function"
 /note="Residues 1 to 76 of 76 are 100.00 pct identical to residues 1 to 76 of 76 from Escherichia coli K-12 strain MG1655: B2810"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC57923.1"
 /db_xref="GI:12517291"
 /translation="MMKTAATISACMLTXLSACSGSNVVMHTNDGRTIVSDGKPTDNDGMISYDANGNKQDINRXDKVENVELDQ"
 625..1830
 /gene="24127"
 625..1830
 /gene="24127"
 /function="orf: Unknown function"
 /note="Residues 1 to 401 of 401 are 99.00 pct identical to residues 1 to 401 of 401 from Escherichia coli K-12 strain MG1655: B2810"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC57924.1"
 /db_xref="GI:12517292"
 /translation="MNVNPNAPFOAFQSPALQDAGVYIDSAATATLKPAVVEATROPYSLSAGNVRISQFAEAOURLTAREAREKRVQQLNRPDRTVTRGTESINMAQCTARLPQDPDEILIVSAEHHANLVPWLMVAQDGAQVRLPLNAQRLPVDDLPELITPRSRILALGOMSVNVTGCPDLARAITFAHSAGIVVVDAGAVHFPADVOODIDFVFSCHLYGPTGIVLYGKSELLEAMSPMIGSKNVEYSFGFTTOSAPMKLEAGTPVACVIGLSXALEMLADYDINOAKENSRSRLATLAEDALAKRPFRRFRODSLLAFDRAGVHSDVYTLAEYGIARAGQHCAPLAEVLGVTGLIRASFAYNRKSDVDAIVNAVDRALLELVLD"
 1830..2273
 /gene="24128"
 1830..2273
 /gene="24128"
 /function="orf: Unknown function"
 /note="Residues 1 to 147 of 147 are 100.00 pct identical to residues 1 to 147 of 147 from Escherichia coli K-12 strain MG1655: B2811"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC57925.1"
 /db_xref="GI:12517293"
 /translation="MTNPOFGAFHFGFTVTAETLRNTEFAPLTOWEDKYROLIMGKOLPALPDELAKAOKETACGECNRYWMLGYTAENCKHMFPCDDSGRIYRGILLAVLTAVECKTAELQKOSPLALDELGLRAQOLSASRSGCLNALSEAIITATKOV"
 complement(2324..3130)
 /gene="24129"
 complement(2324..3130)
 /gene="24129"
 /function="putative enzyme; Not classified"
 /note="Residues 1 to 268 of 268 are 98.88 pct identical to residues 1 to 268 of 268 from Escherichia coli K-12 strain MG1655: B2812"
 /codon_start=1

CDS
 gene
 CDS


```

OY 961 TTCCGCTCCACAGATTCACGCTGCTGCTTTGATTTTGTGCGCTTACATACGCAT 1020
    |||||||
Db 1585 TTCCGCTCCACAGATTCACGCTGCTGCTTTGATTTTGTGCGCTTACATACGCAT 1644
    |||||||
OY 1021 ATGGGACGCGTGGCGGATACGATTCGCTGCGCGGCGGCGGACGATTCGCGCTAG 1080
    |||||||
Db 1645 ATGGGACGCGTGGCGGATACGATTCGCTGCGCGGCGGCGGACGATTCGCGCTAG 1704
    |||||||
OY 1081 CCGCTACTGCGAGATTAAGCGGTAAACCGCACTGCGCGCTGTTTGGCCCATATAT 1140
    |||||||
Db 1705 CCGCTACTGCGAGATTAAGCGGTAAACCGCACTGCGCGCTGTTTGGCCCATATAT 1764
    |||||||
OY 1141 ACAAGACTGATGTGGATGCGCTGCTGATCCGTTACCCGCGCTGGAATTAATGCTG 1200
    |||||||
Db 1765 ACAAGACTGATGTGGATGCGCTGCTGATCCGTTACCCGCGCTGGAATTAATGCTG 1824
    |||||||
OY 1201 GATTAA 1206
    |||||||
Db 1825 GATTAA 1830

RESULT 12
AC084742/c 251891 bp DNA Linear HTG 20-MAY-2002
LOCUS Mus musculus chromosome 2 clone RP23-147D3 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 35 unordered pieces.
AC084742
AC084742.10 GI:20986437
VERSION HTG: HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 251891)
Goltz,J., Halder,A., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
Petere,A., Shim,C., Thomas,E. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 251891)
Goltz,J., Halder,A., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
Petere,A., Shim,C., Thomas,E. and Kucherlapati,R.
Direct Submission
Submitted (11-NOV-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
3 (bases 1 to 251891)
Goltz,J., Halder,A., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
Goltz,J., Halder,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
Petere,A., Shim,C., Thomas,E. and Kucherlapati,R.
Direct Submission
Submitted (20-MAY-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA
On May 20, 2002 this sequence version replaced g1:20219073.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcg.org/Sequence/mouse.html
Contact: hpcg@mednet.harvard.edu
-----Summary Statistics
Center project name: ADK
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 242636 at least Q20
*Consensus quality: 237813 at least Q30
*Estimated insert size: 232938 at least Q40
*Estimated insert size: 251211 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 7.1 x ln Q20 bases; sum-of-contigs estimation

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 87943: contig of 87943 bp in length
* 87944: gap of unknown length
* 87964: contig of 6181 bp in length
* 149804: gap of unknown length
* 149824: gap of unknown length
* 149825: contig of 26177 bp in length
* 176002: gap of unknown length
* 176022: contig of 19195 bp in length
* 195217: gap of unknown length
* 195237: contig of 3627 bp in length
* 198863: gap of unknown length
* 198864: gap of unknown length
* 198883: contig of 881 bp in length
* 199764: gap of unknown length
* 199785: gap of unknown length
* 199785: contig of 1779 bp in length
* 201563: gap of unknown length
* 201564: gap of unknown length
* 201584: contig of 2526 bp in length
* 204110: gap of unknown length
* 204129: gap of unknown length
* 206918: contig of 2790 bp in length
* 206920: gap of unknown length
* 206939: gap of unknown length
* 209157: contig of 2218 bp in length
* 209158: gap of unknown length
* 209177: gap of unknown length
* 209178: contig of 2798 bp in length
* 211975: gap of unknown length
* 211976: gap of unknown length
* 213476: contig of 1481 bp in length
* 213496: gap of unknown length
* 213497: gap of unknown length
* 213497: contig of 1781 bp in length
* 215278: gap of unknown length
* 215298: contig of 1992 bp in length
* 217290: gap of unknown length
* 217310: gap of unknown length
* 219572: contig of 2262 bp in length
* 219572: gap of unknown length
* 219591: gap of unknown length
* 221254: contig of 1663 bp in length
* 221255: gap of unknown length
* 221274: gap of unknown length
* 221275: gap of unknown length
* 221275: contig of 2015 bp in length
* 223280: gap of unknown length
* 223309: gap of unknown length
* 223310: gap of unknown length
* 225551: contig of 2242 bp in length
* 225571: gap of unknown length
* 225572: gap of unknown length
* 225572: contig of 1073 bp in length
* 226644: gap of unknown length
* 226644: gap of unknown length
* 228257: contig of 1593 bp in length
* 228258: gap of unknown length
* 228277: gap of unknown length
* 228278: gap of unknown length
* 230117: contig of 1839 bp in length
* 230136: gap of unknown length
* 230137: gap of unknown length
* 231840: contig of 1704 bp in length
* 231841: gap of unknown length
* 231860: gap of unknown length
* 233408: contig of 1548 bp in length
* 233409: gap of unknown length
* 233428: gap of unknown length
* 233429: gap of unknown length
* 235021: contig of 1593 bp in length
* 235022: gap of unknown length
* 235022: gap of unknown length
* 235041: gap of unknown length
* 237136: contig of 2095 bp in length
* 237137: gap of unknown length
* 237157: gap of unknown length
* 238662: contig of 1506 bp in length
* 238663: gap of unknown length
* 238683: gap of unknown length
* 238772: gap of unknown length
* 238772: contig of 90 bp in length
* 238773: gap of unknown length
* 238792: gap of unknown length
* 239493: contig of 701 bp in length
* 239493: gap of unknown length
* 239494: gap of unknown length
* 241020: contig of 1507 bp in length
* 241021: gap of unknown length
* 241021: gap of unknown length
* 241041: contig of 2340 bp in length
* 243380: gap of unknown length
* 243400: gap of unknown length
* 243401: gap of unknown length
* 243401: contig of 1628 bp in length
* 245029: gap of unknown length
* 245048: gap of unknown length
* 245049: contig of 1410 bp in length
* 246478: gap of unknown length

```

FEATURES

```
* 246479 247886: config of 1408 bp in length
* 247887 247906: gap of unknown length
* 247907 249981: config of 2075 bp in length
* 249982 250001: gap of unknown length
* 250002 251891: config of 1890 bp in length.
Location/Qualifiers
1. 251891
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-147D3"
/sex="male"
1. 87943
```

```
misc_feature /note="assembly_name:Contig326
clone_end:Spe'
vector_side:right"
87964..149804
/note="assembly_name:Contig325"
149825..176001
/note="assembly_name:Contig324
clone_end:T7
vector_side:left"
176022..195216
/note="assembly_name:Contig323"
195237..198863
/note="assembly_name:Contig322"
198884..199764
/note="assembly_name:Contig321"
199785..201563
/note="assembly_name:Contig320"
201584..204109
/note="assembly_name:Contig319"
204130..206919
/note="assembly_name:Contig318"
206940..209157
/note="assembly_name:Contig317"
209178..211975
/note="assembly_name:Contig316"
211996..213476
/note="assembly_name:Contig315"
213497..215277
/note="assembly_name:Contig314"
215298..217289
/note="assembly_name:Contig313"
217310..219571
/note="assembly_name:Contig312"
219592..221254
/note="assembly_name:Contig311"
221275..223289
/note="assembly_name:Contig310"
223310..225351
/note="assembly_name:Contig309"
225372..226644
/note="assembly_name:Contig308"
226665..228257
/note="assembly_name:Contig307"
228278..230116
/note="assembly_name:Contig306"
230137..231840
/note="assembly_name:Contig305"
231861..233408
/note="assembly_name:Contig304"
233429..235021
/note="assembly_name:Contig303"
235042..237136
/note="assembly_name:Contig302"
237157..238662
/note="assembly_name:Contig301"
238683..238772
/note="assembly_name:Contig300"
238793..239493
/note="assembly_name:Contig299"
```

```
misc_feature 239514..241020
/note="assembly_name:Contig298"
misc_feature 241041..243380
/note="assembly_name:Contig297"
misc_feature 243401..245028

Query Match 77.9% Score 939 DB 2 Length 251891:
Best Local Similarity 100.0% Pred. NO. 2.3e-232:
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

OY 1 ATGACGTTTAAATCCCGCCAGTTTCGCCCACTTCCCGCACATGATCGGGC 60
Db 247435 ATGACGTTTAAATCCCGCCAGTTTCGCCCACTTCCCGCACATGCGGC 247376

OY 61 GTCTATCTGACACGCCCGCGGACCGCGCTTAAACCTGAAACCCGTGTGAAGCCACCA 120
Db 247375 GTCTATCTGACACGCCCGCGGACCGCGCTTAAACCTGAAACCCGTGTGAAGCCACCA 247316

OY 121 CAGTTTACAGTGTGAGCGCGCGGAAAGCTCATGCGAGCCAGTTTGGCGGAAGCCAAACGC 180
Db 247315 CAGTTTACAGTGTGAGCGCGCGGAAAGCTCATGCGAGCCAGTTTGGCGGAAGCCAAACGC 247256

OY 181 CTGACCGCGCGTTATGAGCTGACAGAGAAAGTGCGCAATTACTGAATGACCGGAT 240
Db 247255 CTGACCGCGCGTTATGAGCTGACAGAGAAAGTGCGCAATTACTGAATGACCGGAT 247196

OY 241 GATTAACCTATCTGTGAGACCGCGCGGACACCTGATCCATGCAACATGCTGCGACAAATGC 300
Db 247195 GATTAACCTATCTGTGAGACCGCGCGGACACCTGATCCATGCAACATGCTGCGACAAATGC 247136

OY 301 TATGCGGTCGCGGTGTGCAACCGGCGATGATTAATTTGACGCGTGCGAGAACACAC 360
Db 247135 TATGCGGTCGCGGTGTGCAACCGGCGATGATTAATTTGACGCGTGCGAGAACACAC 247076

OY 361 GCCAACCTGTCCTCGCTGATGTGTCGCCAACAACCTGAGACCACAAAGTGTAATTG 420
Db 247075 GCCAACCTGTCCTCGCTGATGTGTCGCCAACAACCTGAGACCACAAAGTGTAATTG 247016

OY 421 CCCGTTAATCGGACGACGACGCGGAGATGTGATTTGTGCCAACAACGATTAATCCCGCT 480
Db 247015 CCCGTTAATCGGACGACGACGCGGAGATGTGATTTGTGCCAACAACGATTAATCCCGCT 246956

OY 481 AGTCGATTTGCGGTTGGGTCAGATGTGCAAGTTACTGCGGTTGCCGATCTCGG 540
Db 246955 AGTCGATTTGCGGTTGGGTCAGATGTGCAAGTTACTGCGGTTGCCGATCTCGG 246896

OY 541 CGAGCGATTACCTTGTCTCATTCAGCGGATGTGATGTTGATGTCGTCACGGG 600
Db 246895 CGAGCGATTACCTTGTCTCATTCAGCGGATGTGATGTTGATGTCGTCACGGG 246836

OY 601 GCAATGATTTCCCGCGGATGTTACGACAACTGATATTTATGCTTTTTCAGGT 660
Db 246835 GCAATGATTTCCCGCGGATGTTACGACAACTGATATTTATGCTTTTTCAGGT 246776

OY 661 CACAACCTGATGAGCGCGGACAGGATGCGGTGATGATGATTAATGAGAACTGCTGAG 720
Db 246775 CACAACCTGATGAGCGCGGACAGGATGCGGTGATGATGATTAATGAGAACTGCTGAG 246716

OY 721 GCGATGTCGCCCTGCTGCGCGGCGGCGGAAATGTTACGAAGTATTTGACGCGTTC 780
Db 246715 GCGATGTCGCCCTGCTGCGCGGCGGCGGAAATGTTGACGAAGTATTTGACGCGTTC 246656

OY 781 ACGACTCAATCTGCGCGGCGGGAACCTGGAACCTGGAACCGGCAATGCTGCTGCTATA 840
Db 246555 ACGACTCAATCTGCGCGGCGGGAACCTGGAACCTGGAACCGGCAATGCTGCTGCTATA 246596

OY 841 GGATTAAGCGCGCGCTGGAATGCTGCGAGATTAACATTAACAACGCGGAAGCTGC 900
Db 246595 GGATTAAGCGCGCGCTGGAATGCTGCGAGATTAACATTAACAACGCGGAAGCTGC 246536

OY 901 AGCGGTAGCTTAGCAAGCTGCGGGAAGATGCGTGGCG 939
Db 246535 AGCGGTAGCTTAGCAAGCTGCGGGAAGATGCGTGGCG 246497
```


LYFFAIPILVSVATNTVESFLTHQALNPPRAILSLILIVGMWTVLRFGEOM
IVKAMSLVFPVALMLLALVLPOMNGAALETSPDASAATGNGLMWTLAIPVM
VSEFNHSPITSSFAVAKREYGEAEKCKSKILARAHIMWTVWEPFVESCULSLPA
DLAAKEDNISLSTLANHPNAPITAMAPITAMAIKSPFGLHGLARBEGRNGVYK
SLRGKKSIEINKKLTALFELMTVATNPISLGLIETLGGPILAMILFLMPMY
AIOKVPARKYSGHISNVFVIMGLAISALFYSLS"

gene
4980.6358
/gene="sdab"
/note="STM2971"
4980.4985
/gene="sdab"
/note="putative RBS for sdab: RegulonDB:STMS1H003023"
4991.6358
/gene="sdab"
/EC_number="4.2.1.13"
/note="similar to E. coli L-serine dehydratase
(deaminase), L-SD2 (AAC75839.1): Blastp hit to AAC75839.1
(455 aa), 94% identity in aa 1 - 455; L-threonine
deaminase 2"
/codon_start=-1
/transl_table=-1
/product="L-serine dehydratase"
/protein_id="AAL21850.1"
/db_xref="GI:16421522"
/translation="MISVFDFKIGIGPSSSHTVGPMKAGQFTDLLRIHLTDVTR
VVDVYGLSLTGKHTDIAIIMGLAGLPTVDIAPFIDVNTGRLNLNGO
HEVERPVKCKMFPADNLSTHNGRITLADGKYLQSPVYSLGGFVDEHSGIT
NSEPVNPIPKYKTAADLORHCOCETLSGLMKQELAHSHLEDFHARFVRENS
GIERGITTEGVLPGLRVPRRAALRMLVSDGDKTTPMAVDLIMKPLAVNENA
AGGRVYATPTNAGCIVPAVLAYDKFIREVANSLARMLVAISAIGSLXKMASISG
AEVGCGEVAGNCSMAAGLALGLSGPRAVOCIAEIAEMHNLGLTCDPAQOVOPC
IERNALSVKAVNAARMAVRITSEPRVCLDKVILEMYETGKDMNKYRETSGLAMK
IYTCO"

gene
6449.7284
/gene="exo"
/note="STM2972"
6449.6454
/gene="exo"
/note="putative RBS for exo: RegulonDB:STMS1H003024"
6469.7284
/gene="exo"
/EC_number="3.1.11.1"
/note="similar to E. coli 5'-3' exonuclease (AAC75840.1):
Blastp hit to AAC75840.1 (281 aa), 87% identity in aa 18 -
281"
/codon_start=-1
/transl_table=-1
/product="exonuclease IX, 5'-3' exonuclease"
/protein_id="AAL21851.1"
/db_xref="GI:16421523"
/translation="MASPRRCYPTRRVRENTVAHLIYDALNLIRIHAVOGSP
CVETCOHALDLIHSOPTHAVAFDDDKRSSGMRHQLPDYKAGRPMPDLNEMIP
ALRAAFEOGRVCMASDGNEDADLAATLAKYTEGHQHTIVSTDGXYCOLLSPGRI
RDYFQKRLMDAPFIEKEFVLPRLPDYGLAGISSKVPAGVIGISPKATOLLIOFO
NLEGYIAHDEVPPEKMRKLETHKEWAFCLRDIAIRLOTDLHIDNLOQLRAR"
complement(7376..8537)
/gene="fucO"
/note="STM2973"

gene

Query Match 69.1% Score 833.2; DB 1: Length 20653;
Best Local Similarity 80.7% Pred. No. 5,3e-205;
Matches 973; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 1 ATGAACGTTTAAATCCCGGAGTTTCGGCCAGTTTCCGCACTACAGATGCGGCG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 18828 ATGAACGTTTAAATCCCGGAGTTTCGGCCAGTTTCCGCGTAGCGGATGCGG 18887
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 61 GTCTATCTGACAGCGCGGACCGGCTTAAACCTGAAGCCTGGTTGAAGCACCCAA 120
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 18888 GTTATCTCGATAGCGCGGACCGGATTAAGCCACAGGCACTCATCTAGCGCAC 18947
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 CAGTTTACAGTCGACGCGGGAAGCGTCATCGAGCGATTGGCGGAAGCCCAACGC 180
||||||| || ||||||||| ||||| ||||| ||||||||| ||||| ||||| |||||

DB 18948 CAGTTTATTTGTTGAGCCCGCGTAACGTTTCATCTGAGCCAGTTTCCGACGCGCAGCC 19007
QY 181 CTGACCGCGCGTTATGACCTGACGAGAGAAAGTGGCGCAATTACTGATGACCGCGAT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19008 CTGACCGCGCGTAATGATAGCGGCGCAGAGCAAAAGCAGCGGCTGTTAAACGCGCCGAT 19067
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 GATTAACATGCTGTCGACGCGCGCGCACCACTGAATCCATGACATGCTGCGCACATGC 300
|| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 19068 GAAAAAGATGCTGCTGACACACCGCGCACACCGAAGCATGACATGCTGCGCGCATGT 19127
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 301 TATGCGGCTCGCGCTGTCGACCGCGCGATGAGATTATGCTGAGCTGCGACAGAACAC 360
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19128 TAGCGCGCTGCTGCTGCGCGCGCGCGATGAATATGTTAGCTGCCCGCGAGCATAC 19187
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 361 GCCAACCTGCTCGCGCTGATGCTGCGCGCGCAACCAACGACCGCAACGATGCAATTC 420
DB 19188 GCCAACCTGCTGCTGCTGATGCTGCGCGCGCAACCAACGCGCGCGATGATTAACATTC 19247
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 421 CCGCTTAATGCGGACGATGCGCGGATGTTGATTTGCTCCAGAACATGATTAATCCCGGT 480
DB 19248 CCGCTTAATGCGGACGATGCGCGGATGTTGATTTGCTCCAGAACATGATTAATCCCGGT 19307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 481 AGTCGATTCCTGCGCTGCGGTCAGATGTCGACGTTACTGCGGCTGCCGATCTGCGG 540
DB 19308 AGTCGATTCCTGCGGTCAGATGTCGACGTTACTGCGGCTGCCGATCTGCGG 19367
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 541 CGAGCATTAACCTTTGCTCATTCAGCGCGGATGCTGATGCTGATGCTGCTGCGG 600
DB 19368 AGCGCATTAACCTTTGCTCATTCAGCGCGGATGCTGATGCTGATGCTGCTGCGG 19427
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 601 GCAGTCATTTCCCGCGCGATGTTACGAACTGATGATTTGATCTTTTCAGGT 660
DB 19428 GCAGTCATTTCCCGCGCGATGTTACGAACTGATGATTTGATCTTTTCAGGT 19487
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 661 CACAACTGTATGCGCGGACAGATGCTGCTGATGCTGATGCTGATGCTGCTGCTGAG 720
DB 19488 CACAACTGTATGCGCGGACAGATGCTGCTGATGCTGATGCTGATGCTGCTGAG 19547
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 721 GCGATGCGCGCTGCGGCGCGCGCGGCAAAATGCTCAGCAATGATGATTTGCGGCTTC 780
DB 19548 GCGATGCGCGCTGCGGCGCGCGCGGCAAAATGCTCAGCAATGATGATTTGCGGCTTC 19607
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 781 ACAGCTCAATCTGCGCGCGGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 840
DB 19608 ACAGCTCAATCTGCGCGCGGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 19667
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 841 GGATTAAGCGCGGCTGGAATGCTGCGCAATGATTAACAGAGCGGCAAGCTGG 900
DB 19668 GGATTAAGCGCGGCTGGAATGCTGCGCAATGATTAACAGAGCGGCAAGCTGG 19727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 901 AGCGCTAGCTTACAGAGCTGCGGGAAGATGCGCTGCGGCAAGCTGCGGCTTCTGTC 960
DB 19728 AGCGCTAGCTTACAGAGCTGCGGGAAGATGCGCTGCGGCAAGCTGCGGCTTCTGTC 19787
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 961 TTCGCTGCGGATTCAGACCTGCTGCGCTTGGATTTGCTGCGCTTCTGCTTACAGAT 1020
DB 19788 TTCGCTGCGGATTCAGACCTGCTGCGCTTGGATTTGCTGCGCTTCTGCTTACAGAT 19847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1021 ATGATGACGCTGCTGCGGATGAGTATGCTGCGGCGCGGCGGAGCATGCGCTGAC 1080
DB 19848 ATGATGACGCTGCTGCGGATGAGTATGCTGCGGCGCGGCGGAGCATGCGCTGAC 19907
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1081 CCGCTACGCGAGATTAAGCGGTAACCGGACACACTGCGGCGCTTTTGGCCATATAT 1140
DB 19908 CCGCTACGCGAGATTAAGCGGTAACCGGACACACTGCGGCGCTTTTGGCCATATAT 19967
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1141 ACAAGAGTGATGATGCTGCTGCTGGAATGCGGTTGACCGCGGCGGCAATTAATGCTG 1200
DB 19968 ACAAGAGTGATGATGCTGCTGCTGGAATGCGGTTGACCGCGGCGGCAATTAATGCTG 20027
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1201 GATTAA 1206
|||||
DB 20028 GATTAA 20033

RESULT 14
AL627277 230050 bp DNA linear BCF 06-JUN-2002
DEFINITION *Salmonella enterica* serovar Typh1 (*Salmonella typhi*) strain CT18,
complete chromosome; segment 13/20.
AL627277 AL513382
VERSION GI:16504049
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 230050),
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebahle,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltham,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
Perry,C., Quail,M., Rutherford,K., Simmonds,M., Skellon,J.,
Stevens,K., Whitehead,S. and Barrall,B.G.
Complete genome sequence of a multiple drug resistant *Salmonella*
enterica serovar Typh1 CT18
Nature 413 (6858), 848-852 (2001)
JOURNAL
MEDLINE
21534947
PUBMED
11677608
REFERENCE
2 (bases 1 to 230050)
Parkhill,J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the *Salmonella*
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
Location/Qualifiers
1. 230050
/organism="Salmonella enterica subsp. enterica serovar
Typh1"
/strain="CT18"
/db_xref="taxon:90370"
309. 1598
/gene="STY3109"
/note="sdac"
309. 1598
/gene="STY3109"
/note="Fasta hit to Q46940 (409 aa), 40% identity in 416
aa overlap
Fasta hit to TDCC_ECOLI (443 aa), 50% identity in 442 aa
overlap
Orthologue of *E. coli* sdac (SDAC_ECOLI): Fasta hit to
SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
/transl_table=11
/codon_start=1
/product="putative serine transporter"
/protein_id="CAD02796.1"
/db_xref="GI:16504050"
/db_xref="SPTREMBL:082435"
/translation="MKTOTSTIASIDSRAMKRTDTMMGLYGAIGAGVLFPLIN
AGVGKPLIMLILAPMTFAHRCITREVLSCNGEDITEVEHFGIGAKLT
LVFFAIPILIVSAVITNTVESFLYHQAINPPRAISLIIYGMHTIVFGEQM
IVKMSILVFAVAMLMALYLIPQNGALLETSDSASTGNGMLMTLMLAIPV
VFSNHSPIISSFAVAKREVEGCAKCKSLIAFAHIMVLTITMPEFSCVSLTPA
DLAAKQONISILSYLANHNPAPIANMAPIAMATKTSFLGHYLCARCGFNGWIK
SLRCKGKSTENKLNKLTALMLVTTIVATNPISLIGMETIGGLIAMLFLMPMT
AIOKVPAMRKYSGHISNVFVIMGLIAISAFVSLFS"
1657. 3024
/gene="STY3110"
/note="sdab"
gene

CDS
1657. 3024
/gene="STY3110"
/note="Fasta hit to TDCC_ECOLI (454 aa), 73% identity in
454 aa overlap
Fasta hit to SDHL_ECOLI (454 aa), 77% identity in 455 aa
overlap
Orthologue of *E. coli* sdab (SDHL_ECOLI): Fasta hit to
SDHL_ECOLI (455 aa), 94% identity in 455 aa overlap"
/transl_table=11
/codon_start=1
/product="L-serine dehydratase 2 (L-serine deaminase 2)"
/protein_id="CAD02797.1"
/db_xref="GI:16504051"
/db_xref="SPTREMBL:082434"
/translation="MISVPDIFKIGIGSPSSHTVCGPKAKQFTDILARHLTDVTR
VVVDVYGSLSGCKGHNHIDIAIIMGLAGMLPDVDDIDAIPIFDVNTGRLLANQ
HEVEFVDKGNHNAHLSLHNGMRITLAKNKVLSOTYYSIGGFIVDEEFCLT
NSEFVVPVRYTADLCHNCOTGLSLGIMONKYLHAKRELOHFAKWEVRS
GTERGTTGCVLPDKLRVPRRAALIRMTYSDOKTTDDNAVVDYMTNMFALVNEEA
AGRVYATPNGAGTVPVAVLATYDKFTREYVNSLARIHMLVASAIGSLYKNAISG
AEVCGEGEVACSMAGLAELGSPVQYCIABEIHNGILTCDPVAGOVQPC
IERNAIASVKAVMARMLRTRSEPRVCLDKVLETMYETGKDMNAVRETSRGLAMK
IVYCD"
3135. 3950
/gene="STY3111"
/note="exo"
3135. 3950
/gene="STY3111"
/note="Orthologue of *E. coli* exo (EX9_ECOLI): Fasta hit to
EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
/transl_table=11
/codon_start=1
/product="5'-3' exonuclease"
/protein_id="CAD02798.1"
/db_xref="GI:16504052"
/db_xref="SPTREMBL:082433"
/translation="MLASPRRECVPPRRVRENTVAHLIYDALNLIRIRHAYGSP
CVETGHALDOLITHSQPHAVAFDDDLARSSRRRORLPDTRKAGPPRPDLHNEP
ALBAFEOKRCVWASDGHEDDLATLAKVTEAHOATVSTDGVCOLLSPGRT
RDYFORMLDAPETKEFEVLPRLPDVYGLAGISSKVPVAGVIGPKSATDLYOFQ
NLEGIYAHLDVEPEKMKRKLTEHKNAPLCRDYARLOTDLHIDGNLODLVR"
3201. 3947
/gene="STY3111"
/note="Pfam match to entry PF01367 5_3_exonuclease, 5'-3'
exonuclease, score 120.40, E-value 3.3e-32"
complement(4042. 5190)
/gene="STY3112"
/note="fucO"
complement(4042. 5190)
/gene="STY3112"
/note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
372 aa overlap
Fasta hit to ADH2_ECOLI (383 aa), 44% identity in 371 aa
overlap
Orthologue of *E. coli* fucO (FUCO_ECOLI): Fasta hit to
FUCO_ECOLI (383 aa), 91% identity in 382 aa overlap"
/transl_table=11
/codon_start=1
/product="1,2-propanediol oxidoreductase (lactaldehyde
reductase)"
/protein_id="CAD02799.1"
/db_xref="GI:16504053"
/db_xref="SPTREMBL:082432"
/translation="MNMRIKLTETAFNGRGAVDALTDEVTRGCVKHALIVTDKTLVOC
GVDDYTSRDADAQWELIAGVITPPTISVVOEGKTFPOSGADVLIAGGSPDPT
KRAVGISINPEPADVRSGLSPTRKPSVPIVAIPTTAALEVTINVTDEEKR
KVCVDPHDIPOAPADIDADMCPALKAAGVVDLTHAIEGYITRAAMALDLHLI
KAETIYAGALRGAVAGEKEGEMALGOVYAGAGCSNUGLIVHCAHPLGAFYRNP
GVANAILLPHVMPFNAGSTNEKFRDIAKAGVIVEGLSDEAFNAVEAVFTLNRDVG
IPHLADVGRKDDIPALAAQAAVDVCTGCPNEASLADIVELYHRAW"
complement(4069. 5169)
/gene="STY3112"
/note="Pfam match to entry PF00465 Fe-ADH, Iron-containing

```
misc-feature      alcohol dehydrogenases, score 655.70, E-value 2.4e-193
                    complement(4354, .4416)
                    /gene="SFY3112"
                    /note="PS00060 Iron-containing alcohol dehydrogenases
signature 2"
                    complement(4585, .4671)
                    /gene="SFY3112"
                    /note="PS00913 Iron-containing alcohol dehydrogenases
signature 1"
                    complement(5207, .5854)
                    /gene="SFY3113"
                    /note="fucA"
                    complement(5207, .5854)
                    /gene="SFY3113"
                    /note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in
179 aa overlap
orthologue of E. coli fucA (FUCA_ECOLI); Fasta hit to
FUCA_ECOLI (215 aa), 92% identity in 215 aa overlap"
                    /codon_start=1
                    /transl_table=11
                    /product="fuculose-1-phosphate aldolase"
                    /protein_id="CAD02800.1"
                    /db_xref="GI:16504054"
                    /db_xref="SPTREMBL:Q82431"
                    /translation="MERRLARQIIDTCLMTRLGKLNQSTAGNSVRYOGKMLTPPTG
IPYKLTSHIVFIDADQHOEGKLPSSMRPMAYOTRPANVAVHNAVHTAVS
ILNRPIAHVMAAGNSITPCARYATPRLSDHVAVALKNKATILQHGIIAC
BENLRKALHAEVEVLAOLYSLTALYDPRVLDDEALIAIYLEKRYGLNIEE"
                    complement(5219, .5845)
                    /gene="SFY3113"
                    /note="Pfam match to entry PF00596 Aldolase-II, Class II
Aldolase and Adducin N-terminal domain., score 337.10,
E-value 2e-97"
                    6410..7726
                    /gene="SFY3115"
                    /note="fucP"
                    6410..7726
                    /gene="SFY3115"
                    /note="Orthologue of E. coli fucP (FUCP_ECOLI); Fasta hit
to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
                    /codon_start=1
                    /transl_table=11
                    /product="L-fucose permease"
                    /protein_id="CAD02801.1"
                    /db_xref="GI:16504055"
                    /db_xref="SPTREMBL:Q82430"
                    /translation="WGNNTITQTSFRVDAEDSKSKRYIIPALLCSFFLVAVANNL
NDLLPFOQAFLLTFQAGLQSAFYFQYFVPIPIAGLIMKLSKAGIITGLFLYA
FGAALFMPAAEIMNTYLLFLIGLITIAAGCLETANPVTALGSESGHFRNLQOT
FNSFGAIIAVVFGOSLISNVPHOSEALDKMTPDLSAYKSLVSVOTPYMIIAVI
VLVALLIMLTFFPALSDSDSDAKOSFSLSLRILIRHHRMVAIVLQFVGAOTA
CHSYLRIVIEIPEPTPGFANNTUTGTWVCFIFGFTGTHLISRPAPRYLAALAL
AHLCLISAFSGHIGLVALTLCSAFMSIOTYPTITSLGINKLQODTKKSSFTVMTII
GGGIYTPVWGFSDAAGKIPITAEVLPALCFAVIFIFARRSOAATN"
                    7758..9488
                    /gene="SFY3116"
                    /note="fucI"
                    7758..9488

gene
CDS
Query Match      68.3%; Score 823.6; DB 1; Length 23050;
Best Local Similarity 80.2%; Pred. No. 1.9e-202;
Matches 967; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

OY 1 ATCAAGCTTTTATCCCGCGCAGTTTCCGCCAGTTTCCGCACTACAGATGGGGC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15494 ATGAAGCTTTTATCCCGCGCAGTTTCCGCCAGTTTCCGCACTACAGATGGGGT 15553

OY 61 GTCTATCTGACAGCGCGCGGCGCTTAACGTAACGCGTGGTGAAGCACCCAA 120
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15554 GTTATCTGATAGCGCGCGCGCGCTTAACGTAACGCGTGAAGCACCCAA 15613

OY 121 CAGTTTACAGTCTGAGCGCGCGAAGCTCATCGCAGCGCTTTCGGAAGCCCAACGC 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 15614 CAGTTTATTTGTTGAGCGCGCGGTAACTTCATCGTAGCA6TTCGCGAGCGCGC 15673
OY 181 CTGACCGCGCGTATAGACTGACGAGAGAAAGTGCGCAATTACTGAATGACCGGAT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15674 CTGACCGCGCGTATAGACTGACGAGAGAAAGTGCGCGCTTTAAAGCGCGCGAT 15733

OY 241 GATTAAGTATGCTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15734 GAAAAAGTATGCTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15793

OY 301 TATGCGCGTCCGCGTGTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15794 TACGCGCGTCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15853

OY 361 GCCAACCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15854 GCCAACCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15913

OY 421 CCGCTTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15914 CCGCTTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15973

OY 481 AGTCGATTCGCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15974 AGTCGATTCGCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 16033

OY 541 CGAGCGATTAACGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16034 GCGGCTATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16093

OY 601 GCAGTCATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16094 GCGTACACTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16153

OY 661 CACAACTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16154 CACAACTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16213

OY 721 GCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16214 GCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16273

OY 781 ACAGCTCAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16274 ACAGCTCAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16333

OY 841 GGAATTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16334 GCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16393

OY 901 AGCGGTAGCTTACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16394 AGCGGTAGCTTACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16453

OY 961 TTCGCGTCCGAGATTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16454 TTCGCGTCCGAGATTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16513

OY 1021 ATGCTACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16514 ATGCTACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16573

OY 1081 CCGCTACTGCGAATTAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16574 CCGCTACTGCGAATTAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16633

OY 1141 ACAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16634 ACCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 16693

OY 1201 GATTA 1206
    |||||
DB 16694 GATTA 16699
```

RESULT 15
LOCUS AE013916 9280 bp DNA linear BCT 26-JUL-2002
DEFINITION Yersinia pestis KIM section 316 of 415 of the complete genome.
ACCESSION AE013916 AE009952
VERSION AE013916.1 GI:21960078
KEYWORDS
SOURCE Yersinia pestis KIM.
ORGANISM Yersinia pestis KIM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
REFERENCE 1 (bases 1 to 9280)
AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
TITLE Genome Sequence of Yersinia pestis KIM
JOURNAL J. Bacteriol. 184 (16), 4601-4611 (2002)
PUBMED 12142430
REFERENCE 2 (bases 1 to 9280)
AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE
LOCATION/Qualifiers
1..9280
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
complement(102..1019)
/gene="gcvA"
/note="y3154"
/codon_start=1
/gene="gcvA"
complement(102..1019)
/function="regulator: central intermediary metabolism:
Pool, multipurpose conversions"
/note="residues 1 to 304 of 305 are 88.15 pct identical to
residues 1 to 304 of 305 from E. coli K12 : B2808;
residues 1 to 304 of 305 are 89.47 pct identical to
residues 1 to 304 of 305 from GenPept : >gblAL21860.1
(AE008836) regulator of gcv operon (LysR family)
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="positive regulator of gcv operon"
/protein_id="AA06704.1"
/db_xref="GI:21960078"
/translation="MSKRLPLNALRFAAARLSTKAEEFYTAQAVSHQIKSL
EDFLGLFRNRNSLILTEGGSYIDIEFTSINEATRKLOARAKAGALVSLP
SFAIQLVPLRSGNAVPGIDIVAVREDEKLVADVAIFYGGNSGRTERL
YAEFLVPCAPSLTGTENGCLKPSDLNHNLTLDHSTRDLAVTROLGVPOINVOQGP
IFSHSAFVVOAAVGGIGALVNNVMAQSEIEARLVCFFNDVLVSKNAFYLVCHDSOA
ELCKIAFRWMIARASAEQKLRFRREN"
1510..1692
/gene="y3155"
1510..1692
/gene="y3155"
/function="unknown"
/note="residues 12 to 60 of 60 are 35.08 pct identical to
residues 559 to 613 of 752 from GenPept :
>gblAK76896.1(AE001438-149 (AE001438) Integrin-like
repeats domain fused to lysozyme, LysC 91ycosyl hydrolase
[Clostridium acetobutylicum]"
/codon_start=1
/transl_table=11
/product="hypothetical"

gene
CDS
/protein_id="AA06705.1"
/db_xref="GI:21960080"
/translation="MTDHLVLIYFRFTISNDSATITNNVNMHNI IAOIRDHKSK
KISGSLTCCCOOYESF"
1679..2884
/gene="y3156"
1679..2884
/gene="y3156"
/note="residues 1 to 401 of 401 are 67.08 pct identical to
residues 1 to 401 of 401 from E. coli K12 : B2810"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA06706.1"
/db_xref="GI:21960081"
/translation="MKVFNMDFRREPALSDKLTYSLSAATLAKPRAMIDATQFQ
QDSATVYSHOSALSTVFEENTROVADFINSTAEINITRTGTEAINLAOSYA
RPLQREDETVSEAEHNLIPMLVARDGTQIKYKPLGLDHLIDLOLPOLNLEK
TRLIAGQSNVNTGSGPDLAQRILRHOQGVYVYVAGOCIVTPADVOALDIFYAF
SSHKLGPFGIGVIGKTELEEMPAHQGKMLTHASFGCFPHBVPKPEAGTQNI
AGVIGLSAVLKMLEHIDLEAEVYSGLATMAENKLAOLPGFHSYRCCOOSLAFPF
GVHSDLVALLAEQIALRAGQCAQPLMAALGVNCSLRASFAPVNTPDVEMLCAL
GKALELRD"
3019..3462
/gene="y3157"
3019..3462
/gene="y3157"
/note="residues 4 to 137 of 147 are 58.95 pct identical to
residues 7 to 140 of 147 from E. coli K12 : B2811"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA06707.1"
/db_xref="GI:21960082"
/translation="MMIAPFPGHDITATDLIEKPSAHKOMEDRYROLILAKOLPPL
QEAWKNELELTCGNCENVMWGLHPLDGTLEHFGSEGRVRLVILTVAGSKTPQ
OVLADPLALFEOLGLROCLSTSRANGLOALAGCVTIKAKYER"
complement(3498..4325)
/gene="y3158"
complement(3498..4325)
/gene="y3158"
complement(3498..4325)
/function="putative"
/note="residues 1 to 267 of 275 are 80.52 pct identical to
residues 1 to 267 of 268 from E. coli K12 : B2812;
residues 1 to 267 of 275 are 80.89 pct identical to
residues 1 to 267 of 268 from GenPept : >gblAL21865.1
(AE008837) paral putative enzyme [Salmonella typhimurium
LT2]"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AA06708.1"
/db_xref="GI:21960083"
/translation="MSTAYSEAYORFGIARLYGGOALAFSAHVCVIGTCGSM
AARALRTGICATITLDMDDVCTNNRNRGRIHALRHNIQAKTEVMAERLAIPECHV
TCIDDTTADANVALELNKNSVYIDAISVRPAALLSTCRRIKIPVTTGACGQID
PRTIAVDLAKTITQDPLAALRLRRLKSDRVNYSRKGKIGIDCFSSSEPLVYPOAGS
VCASRTAEKPKMDCSTGCSATMTATGFAVYSHVLKMKMAKAAROIHESVATAS"
complement(4468..5640)
/gene="y3159"
complement(4468..5640)
/gene="y3159"
complement(4468..5640)
/gene="y3159"
complement(4468..5640)
/function="enzyme; murein sacculus, peptidoglycan"
/note="residues 1 to 366 of 390 are 69.94 pct identical to
residues 1 to 364 of 365 from E. coli K12 : B2813"
/codon_start=1
/transl_table=11
/product="membrane-bound lytic murein transglycosylase A"
/protein_id="AA06709.1"
/db_xref="GI:21960084"
/translation="MTRMGKYLKSGIMAVIAGCSGPTDRGQYKDGDELSLEY
NEPNANGKPYNAKDYSDQYKVINQSSPGLYNRSDTFNVAONMLAGADTSKLSLFLG

Tue Apr 22 16:31:44 2003

us-09-622-419-3.rge

Page 24

Search completed: April 22, 2003, 12:23:01
Job time : 3767.45 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 08:16:32 ; Search time 302.905 Seconds

(Without alignments)
8966.219 Million cell updates/sec

Title: US-09-622-419-3

Perfect score: 1206
Sequence: 1 atgaacgttttaaccgcgc.....tggaattatgttgattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1206	100.0	1206	AA209785	E. coli biosl DNA.
2	1206	100.0	20	AA209785	DE19731274 Seq ID
3	1206	100.0	3794	AA209791	Plasmid pHS1 biosl
4	1206	100.0	3794	AA209792	DE19731274 Seq ID
5	1206	100.0	4975	AA209792	Plasmid pHS1 metk/
6	1206	99.9	1216	AA209792	E. coli biosl ORF
7	369	30.6	524	ABQ48374	Oligonucleotide fo
8	369	30.6	524	ABQ48375	Oligonucleotide fo
9	348	28.9	524	ABQ48376	Oligonucleotide fo

Result	Score	Query Match	Length	DB ID	Description
10	348	28.9	524	ABQ48377	Oligonucleotide fo
11	287.6	23.8	917	ABQ23332	Oligonucleotide fo
12	287.6	23.8	917	ABQ23333	Oligonucleotide fo
13	277	23.0	917	ABQ23330	Oligonucleotide fo
14	277	23.0	917	ABQ23331	Oligonucleotide fo
15	244	20.2	1221	AA209787	E. coli biosl DNA.
16	234.6	19.5	1217	ABK77448	Bacillus clausii g
17	234.6	19.5	1224	ABK77447	Bacillus clausii g
18	225	18.7	4160	ABQ71003	Listeria monocytog
19	215.4	17.9	684707	ABQ67196	Listeria innocua C
20	215.4	17.9	3011208	ABQ69245	Listeria innocua D
21	207.4	17.2	2944528	ABA03041	Listeria monocytog
22	204.8	17.0	1830121	AAT42063	Haemophilus influenzae
23	198.4	16.5	10397	AA133227	Enterococcus faec
24	192.4	16.0	18627	AAV52246	Streptococcus pneu
25	190.6	15.8	1035	ABQ68818	Listeria monocytog
26	175.2	14.5	4280	AAV74392	Staphylococcus aur
27	170.2	14.1	598	AA584162	DNA encoding novel
28	165.2	13.7	2365589	ABA50521	Genomic sequence o
29	160	13.3	1224	ABQ68933	Streptococcus poly
30	153.8	12.8	1230	ABQ68932	Streptococcus poly
31	153.8	12.8	2135361	ABN71527	Streptococcus poly
32	153.4	12.7	1230025	AAV91990	Nucleotide sequenc
33	151.2	12.5	1287	ABN92579	Staphylococcus epi
34	148	12.3	4973	AAV58289	S. pyogenes SP-55-
35	145.2	12.0	727	ABK73013	Bacillus lichenifo
36	138	11.4	265	AA551198	Salmonella typhimu
37	129	10.7	13256	AAV20528	Propionibacterium
38	125.4	10.4	21719	AA559568	Propionibacterium
39	119.4	9.9	1263	AA66592	C. glutamicum codin
40	119.4	9.9	1386	AAV72081	Corynebacterium gl
41	119.4	9.9	349680	AAV68529	C. glutamicum codin
42	99.2	8.2	534	ABN78987	Human isomerase-11
43	90.4	7.5	761	AAT25823	C. glutamicum nltS
44	87.6	7.3	349980	AAT6431	Pyrococcus abyssal
45	74	6.1	4443	AA571615	DNA encoding novel

ALIGNMENTS

RESULT 1
ID AA209785 standard; DNA: 1206 BP.
XX
AC AA209785:
XX
DT 23-NOV-1999 (first entry)
XX
DE E. coli biosl DNA.
XX
XX SAM: S-adenosyl methionine synthase; biosynthesis; biotin; biosl; bios2;
KM bios3; cofactor; decarboxylation; Vitamin H; ss
XX
OS Escherichia coli.
XX
FH Key
FH Location/Qualifiers
FT CDS
FT 1..1206
FT /*tag= a
FT /*product= "biosl"
XX
PN DE19806872-A1.
XX
XX 26-AUG-1999.
XX
XX 19-FEB-1998: 98DE-1006872.
XX
XX 19-FEB-1998: 98DE-1006872.
XX
XX (BADI) BASF AG.
XX
XX Schroeder H;
XX

DR WPI: 1999-480095/41.
DR P-PSDB: AAV33264.
PT Production of biotin by expressing S-adenosyl-methionine synthase and
PT second biotin synthesis gene in host cells
XX
XX
PS Claim 1: Page 12-14; 48pp; German.
XX
CC This invention describes a novel method for the preparation of biotin
CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
CC sequence (1), and (b) at least one of the other biotin biosynthesis
CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC most animals and some microorganisms. Expression of biotin plus bios1,
CC bios2 or bios3 significantly increases productivity of biotin
CC biosynthesis, particularly by at least 3 times. This sequence encodes
CC the Escherichia coli bios1 protein which is used in the method of the
CC invention.
XX
XX Sequence 1206 BP; 256 A; 315 C; 359 G; 276 T; 0 other:
SO
Query Match 100.0%; Score 1206; DB 20; Length 1206;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1206: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAACGTTTTTAAATCCCGCGAGTTTCGCCGCAAGTTTCCGCACTACAGATCGCGC 60
DB 1 ATGAACGTTTTTAAATCCCGCGAGTTTCGCCGCAAGTTTCCGCACTACAGATCGCGC 60
OY 61 GTCCTATCTGCAGACGGCGCGGACCGGCTTAACCTGAAGCGGTGGTGAACCCCAA 120
DB 61 GTCCTATCTGCAGACGGCGCGGACCGGCTTAACCTGAAGCGGTGGTGAACCCCAA 120
OY 121 CAGTTTACAGTCTGAGCGCGGAAACGTCATCCAGCAGCTTGGCCGAACCCCAACGC 180
DB 121 CAGTTTACAGTCTGAGCGCGGAAACGTCATCCAGCAGCTTGGCCGAACCCCAACGC 180
OY 181 CTGACCGCGCGCTTATGAACTGACGAGAGAAAGTGGCGCAATTAAGTACACCGAT 240
DB 181 CTGACCGCGCGCTTATGAACTGACGAGAGAAAGTGGCGCAATTAAGTACACCGAT 240
OY 241 GATTAACATGATGCTGAGCGCGGACCACTGAATCCATCAATAGTGTGCACAATGC 300
DB 241 GATTAACATGATGCTGAGCGCGGACCACTGAATCCATCAATAGTGTGCACAATGC 300
OY 301 TATCGCGCTCCGCTGCTGCAACCGGCGATGATTAATGTCAGCGTGCAGAACACGC 360
DB 301 TATCGCGCTCCGCTGCTGCAACCGGCGATGATTAATGTCAGCGTGCAGAACACGC 360
OY 361 GCCAACCTGCTGCCGTGATGTCGCCCAACAACCTGAGCCAAAGTGTGAATTTG 420
DB 361 GCCAACCTGCTGCCGTGATGTCGCCCAACAACCTGAGCCAAAGTGTGAATTTG 420
OY 421 CGGCTTAATGCGCAGACTGCGGATGTCATTGTTGCCAGACTGATTAATCCCGCT 480
DB 421 CGGCTTAATGCGCAGACTGCGGATGTCATTGTTGCCAGACTGATTAATCCCGCT 480
OY 481 ACTCGGATTTCTGGCTGGTCAGATGTCGAACGTTACTGCGGTTGCCGCGATCTGGC 540
DB 481 ACTCGGATTTCTGGCTGGTCAGATGTCGAACGTTACTGCGGTTGCCGCGATCTGGC 540
OY 541 CGAGCGATTACTTTGCTCATTCACCGCGGATGTCGATGATGTCCTCAGGGG 600
DB 541 CGAGCGATTACTTTGCTCATTCACCGCGGATGTCGATGATGTCCTCAGGGG 600
OY 601 GCAGTGCATTTCCCGGGGATGTCAGCAACGATGATTTCTATGCTTTTCAGGT 660
DB 601 GCAGTGCATTTCCCGGGGATGTCAGCAACGATGATTTCTATGCTTTTCAGGT 660
OY 661 CACAACCTGTATGCGCGCAGACTGTCGCTGTATGTAATCAACACTGCTGGAG 720
DB 661 CACAACCTGTATGCGCGCAGACTGTCGCTGTATGTAATCAACACTGCTGGAG 720

OY 721 CGCATGTCGCCCTGCTGGCGCGGCAAAATGCTTACGAAATGTTTGGACGCTTC 780
DB 721 CGCATGTCGCCCTGCTGGCGCGGCAAAATGCTTACGAAATGTTTGGACGCTTC 780
OY 781 ACGACTCAATCTGCGCGCTGGAACACTGGAACCTGGAACCGCAAAATGCTGCTCTA 840
DB 781 ACGACTCAATCTGCGCGCTGGAACACTGGAACCTGGAACCGCAAAATGCTGCTCTA 840
OY 841 GGATTAACCGCGCGCTGGAATGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCT 900
DB 841 GGATTAACCGCGCGCTGGAATGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCT 900
OY 901 AGCCGTACTAGCAACCGCTGCGGAGATGCGCTGCGGAAACGTCGCCGCTTTCGTTCA 960
DB 901 AGCCGTACTAGCAACCGCTGCGGAGATGCGCTGCGGAAACGTCGCCGCTTTCGTTCA 960
OY 961 TTCGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAG 1020
DB 961 TTCGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAG 1020
OY 1021 ATGCTGAGCTGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGAT 1080
DB 1021 ATGCTGAGCTGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGAT 1080
OY 1081 CCGCTACTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAG 1140
DB 1081 CCGCTACTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAGATGCTGCGGAG 1140
OY 1141 ACAAGAGTGTATGTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 ACAAGAGTGTATGTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1201 GATTAA 1206
DB 1201 GATTAA 1206

RESULT 2
AXX02814
ID AXX02814 standard; DNA; 3465 BP.
XX
XX AXX02814:
AC
XX
XX 14-MAY-1999 (first entry)
XX
DE19731274 Seq ID 9.
XX
XX Biotin: synthesis: dehydrobiotin; yield increase: industry:
XX
XX Fermentation, ss.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH 272,1477
FT CDS /'tag- a
FT
XX
XX DE19731274-A1.
XX
XX 28-JAN-1999.
XX
XX 22-JUL-1997; 97DE-1031274.
XX
XX 22-JUL-1997; 97DE-1031274.
XX
XX (BADI) BASF AG.
XX
XX Hauer B, Schroeder H:
XX
XX WPI: 1999-107030/10.
XX
XX P-PSDB: AAW92938.
XX
XX Improved synthesis of biotin by expressing the bios1 or bios2

PT sequence in biotin-producing cells - and related gene constructs.
 PT provides increased conversion of dehydrobiotin in eukaryotic or
 PT prokaryotic hosts

PS Example 2; Page 32-35; 48pp; German.

XX This invention describes a method for the synthesis of biotin in
 CC Escherichia coli. This method involves the expression of a biotin gene
 CC or its functional variants, analogues and derivatives, in a prokaryote
 CC or eukaryote that is able to produce biotin. The cells are grown and
 CC the biotin produced either used directly, after separation of the
 CC biomass, or after purification. Constructs containing this nucleic acid
 CC or protein or their variants etc., can be coupled to one or more
 CC regulators for increasing gene and/or protein expression, and/or having
 CC its natural regulators 'switched off'. Expression of this biotin protein
 CC leads to increased conversion, by at least 3-fold, of dehydrobiotin to
 CC biotin, thus increasing yield and making possible an industrially useful
 CC fermentative method for biotin production.

SO Sequence 3465 BP; 804 A; 963 C; 934 G; 764 T; 0 other:

Query Match 100.0%; Score 1206; DB 20; Length 3465;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACGTTTATATCCCGCAGTTTCCGCGCCAGTTTCCCGCAGTACAGATGCGGGC 60
 DB 272 ATGACGTTTATATCCCGCAGTTTCCGCGCCAGTTTCCCGCAGTACAGATGCGGGC 331
 OY 61 GTCTATCTGACAGCGCGCGACCGGCTTAAACCTGAAGCCGTGTTGAAGCCACCAA 120
 DB 332 GTCTATCTGACAGCGCGCGACCGGCTTAAACCTGAAGCCGTGTTGAAGCCACCAA 391
 OY 121 CAGTTTACAGTGAAGCGCGGAAAGCTGCAGCAGCCAGTTTCCGCGCGGCGGCGGCGG 180
 DB 392 CAGTTTACAGTGAAGCGCGGAAAGCTGCAGCAGCCAGTTTCCGCGCGGCGGCGGCGG 451
 OY 181 CTGACCGCGGTTATAGCTGACGAGAGAAAGTGGCGCAATTAATGACCGGAT 240
 DB 452 CTGACCGCGGTTATAGCTGACGAGAGAAAGTGGCGCAATTAATGACCGGAT 511
 OY 241 GATMAAACTATGCTGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 DB 512 GATMAAACTATGCTGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
 OY 301 TATGCCCGTCCGCTGCTGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 DB 572 TATGCCCGTCCGCTGCTGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
 OY 361 GCCACCTGCTCCGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 DB 632 GCCACCTGCTCCGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
 OY 421 CCGCTTAATGCGGAGCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 DB 692 CCGCTTAATGCGGAGCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751
 OY 481 AGTCGATTTGCGGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 DB 752 AGTCGATTTGCGGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811
 OY 541 CGAGCGGTTACGTTGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 DB 812 CGAGCGGTTACGTTGCTGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 871
 OY 601 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATGATTTCTATGCTTTTCAGGT 660
 DB 872 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATGATTTCTATGCTTTTCAGGT 931
 OY 661 CACAACTGATGCGCGGAGAGTATGCGGCTGATGCTGATGCTGATGCTGATGCTGAG 720
 DB 932 CACAACTGATGCGCGGAGAGTATGCGGCTGATGCTGATGCTGATGCTGATGCTGAG 991

OY 721 GCGATGTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
 DB 992 GCGATGTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1051
 OY 781 ACGACTCATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 DB 1052 ACGACTCATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1111
 OY 841 GCGATGTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 DB 1112 GCGATGTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1171
 OY 901 AGCGTACGTTAGCAACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 DB 1172 AGCGTACGTTAGCAACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1231
 OY 961 TTCGCTGCGGAGATTCACAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 DB 1232 TTCGCTGCGGAGATTCACAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291
 OY 1021 ATGCTACGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 DB 1292 ATGCTACGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1351
 OY 1081 CCGCTACTGCGAGAAATTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 DB 1352 CCGCTACTGCGAGAAATTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1411
 OY 1141 ACAAGAGTATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
 DB 1412 ACAAGAGTATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1471
 OY 1201 GATTAA 1206
 DB 1472 GATTAA 1477

RESULT 3
 ID AA209791 standard; DNA; 3794 BP.
 AC AA209791:
 XX 23-NOV-1999 (first entry)
 DT XX
 DE Plasmid pHS1 bios1 DNA.
 XX
 KW SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bios1; bios2;
 KM bios3; cofactor; decarboxylation; Vitamin H; ss.
 XX
 OS Synthetic.
 XX
 FH key Location/Qualifiers
 FT CDS 601..1806
 FT /tag= a
 FT /product= "bios1"
 XX
 PN DE19806872-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 19-FEB-1998; 98DE-1006872.
 XX
 PR 19-FEB-1998; 98DE-1006872.
 XX
 PA (BADI) BASF AG.
 XX
 PI Schroeder H.
 XX
 DR WPI: 1999-480095/41.
 DR P-PSDB: AAY33268.
 XX
 PT Production of biotin by expressing S-adenosyl-methionine synthase and

PT second biotin synthesis gene in host cells -

XX
PS Example 3: Page 30-33; 48pp: German.

CC This invention describes a novel method for the preparation of biotin
CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
CC sequence (1), and (b) at least one of the other biotin biosynthesis
CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC most animals and some microorganisms. Expression of biotin plus bios1,
CC biosynthesis, particularly by at least 3 times. This sequence encodes
CC the bios1 protein from plasmid pBS1 which is used in the method of the
CC invention.

SO Sequence 3794 BP; 910 A; 1034 C; 1020 G; 830 T; 0 other:

Query Match 100.0%; Score 1206; DB 20; Length 3794;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACGTTTAAATCCCGCGAGTTTCCCGCCAGTTTCCCGCACTACAGATGCGGC 60
DB 602 ATGACGTTTAAATCCCGCGAGTTTCCCGCCAGTTTCCCGCACTACAGATGCGGC 660
OY 61 GTCTATCTGCAGACGCGCGCGACCGCGCTTAAACCTGAAGCCGTGTTGAACCCACCA 120
DB 661 GTCTATCTGCAGACGCGCGCGACCGCGCTTAAACCTGAAGCCGTGTTGAACCCACCA 720
OY 121 CAGTTTACAGTCTGAGGCGCGGAAAGCTCCATCCAGCCAGTTTCCCGCACTACAGATGCGGC 180
DB 721 CAGTTTACAGTCTGAGGCGCGGAAAGCTCCATCCAGCCAGTTTCCCGCACTACAGATGCGGC 780
OY 181 CTGACCGCGCGCTTATGACAGTCTGACAGAGAAAGTGGCGCAATTACTGAAATGCACCGAT 240
DB 781 CTGACCGCGCGCTTATGACAGTCTGACAGAGAAAGTGGCGCAATTACTGAAATGCACCGAT 840
OY 241 GATAAAGTATGCTGTGAGACGCGCGGACCACTGAATCCATCAACATGCTGGCACAATGC 300
DB 841 GATAAAGTATGCTGTGAGACGCGCGGACCACTGAATCCATCAACATGCTGGCACAATGC 900
OY 301 TATGGCGCTCCGCGTCTGCAACCGCGCGATGATATTATTCAGCGTGGCAGAACACAC 360
DB 901 TATGGCGCTCCGCGTCTGCAACCGCGCGATGATATTATTCAGCGTGGCAGAACACAC 960
OY 361 GCCAAGCTGTCGCCCTGCTGATGCTGCCCAACAAGTGGAGCCAAAGTGGTGAATTC 420
DB 961 GCCAAGCTGTCGCCCTGCTGATGCTGCCCAACAAGTGGAGCCAAAGTGGTGAATTC 1020
OY 421 CCGCTTAATGCCGACGACTGCGGATGCTGATTTGTTGCCAGAACTGATTACTCCCGT 480
DB 1021 CCGCTTAATGCCGACGACTGCGGATGCTGATTTGTTGCCAGAACTGATTACTCCCGT 1080
OY 481 AGTGGATTTCGGCGTGGTGCAGATGTCGAACCTTACTGCGGTTGCCCGGATTCGGC 540
DB 1081 AGTGGATTTCGGCGTGGTGCAGATGTCGAACCTTACTGCGGTTGCCCGGATTCGGC 1140
OY 541 CGAGCGATTACTTGTCTCATTCACCGCGGATGCTGATGCTGATGCTGATGCTGATGCTG 600
DB 1141 CGAGCGATTACTTGTCTCATTCACCGCGGATGCTGATGCTGATGCTGATGCTGATGCTG 1200
OY 601 GCAGTGCATTTCCCGCGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 660
DB 1201 GCAGTGCATTTCCCGCGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1260
OY 661 CACAAGCTGATGCGCGCAGATGCTGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 720
DB 1261 CACAAGCTGATGCGCGCAGATGCTGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1320
OY 721 CGGATGTCGCCCTGCTGGCGCGCAAAATGCTTCAAGAGTGGTGGTGGTGGTGGTGGTGGTGG 780
DB 1321 CGGATGTCGCCCTGCTGGCGCGCAAAATGCTTCAAGAGTGGTGGTGGTGGTGGTGGTGGTGG 1380

OY 781 ACAGTCAATCTGCGCGCTGGAAACTGCAAGCTGCAACGCCAAATGCTGCTGTCTATA 840
DB 1381 ACAGTCAATCTGCGCGCTGGAAACTGCAAGCTGCAACGCCAAATGCTGCTGTCTATA 1440
OY 841 GGATTAACCGCGCGCTGGAATGCTGCTGCAATTAACGATTAACACCGCGGAAAGCTGG 900
DB 1441 GGATTAACCGCGCGCTGGAATGCTGCTGCAATTAACGATTAACACCGCGGAAAGCTGG 1500
OY 901 AGCGTACTTACGCAACGCTGGCGGAGATGCGCTGGCGGAAAGCGCCGGCTTTCGTTCA 960
DB 1501 AGCGTACTTACGCAACGCTGGCGGAGATGCGCTGGCGGAAAGCGCCGGCTTTCGTTCA 1560
OY 961 TTCGCTCCAGAGATTCAGCGCTGCTGCTGCTTGAATTTGCTGGCTTATCATAGCGAT 1020
DB 1561 TTCGCTCCAGAGATTCAGCGCTGCTGCTGCTTGAATTTGCTGGCTTATCATAGCGAT 1620
OY 1021 ATGTGACGCTGCTGCGCGGATGCGTATTTGCTGCGCGCGCGCGGCGACATTCGCTAG 1080
DB 1621 ATGTGACGCTGCTGCGCGGATGCGTATTTGCTGCGCGCGCGCGGCGACATTCGCTAG 1680
OY 1081 CCGCTACTGCGCAATTAAGCGCTAACCGCGACACTGCGCGCTCTTTTGGCGCATATAT 1140
DB 1681 CCGCTACTGCGCAATTAAGCGCTAACCGCGACACTGCGCGCTCTTTTGGCGCATATAT 1740
OY 1141 ACAAGAGTGAATGTGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
DB 1741 ACAAGAGTGAATGTGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
OY 1201 GATTAA 1206
DB 1801 GATTAA 1806

RESULT 4
AA02812
ID AA02812 standard; DNA: 3794 BP.
XX
AC AA02812;
XX
DT 14-MAY-1999 (first entry)
XX
DE DE19731274 Seq ID 5.
XX
KW Biotin; synthesis; dehydrobiotin; yield increase; industry;
KW fermentation; ss.
XX
OS Escherichia coli.
XX
FT Key Location/Qualifiers
FT CDS 601..1806
FT /tag= a
FT
PN DE19731274-A1.
XX
PD 28-JAN-1999.
XX
PF 22-JUL-1997; 97DE-1031274.
XX
PR 22-JUL-1997; 97DE-1031274.
XX
PA (BAD1) BASF AG.
XX
PI Hauer B, Schroeder H;
XX
DR WPI: 1999-107030/10.
XX
P-PSDB: AAM92936.
XX
PT Improved synthesis of biotin by expressing the bios1 or bios2
PT sequence in biotin-producing cells - and related gene constructs,
PT provides increased conversion of dehydrobiotin in eukaryotic or
PT prokaryotic hosts

PS Example 2: Page 21-24: 48bp: German.
 XX This invention describes a method for the synthesis of biotin in
 CC Escherichia coli. This method involves the expression of a biotin gene
 CC or its functional variants, analogues and derivatives. In a prokaryote
 CC or eukaryote that is able to produce biotin. The cells are grown and
 CC the biotin produced either used directly, after separation of the
 CC biomass, or after purification. Constructs containing this nucleic acid
 CC or protein or their variants etc., can be coupled to one or more
 CC regulators for increasing gene and/or protein expression, and/or having
 CC its natural regulators switched off. Expression of this biotin protein
 CC leads to increased conversion, by at least 3-fold, of dehydrobiotin to
 CC biotin, thus increasing yield and making possible an industrially useful
 CC fermentative method for biotin production.

XX Sequence 3794 BP; 910 A; 1034 C; 1020 G; 830 T; 0 other;

Query Match 100.0%; Score 1206; DB 20; Length 3794;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAACGTTTATATCCCGGACGTTTCCGCCACTATACAGATGCGGC 60
 DB 601 ATGAACGTTTATATCCCGGACGTTTCCGCCACTATACAGATGCGGC 660
 OY 61 GTCTATCTGCACGCGCGGACGCGCTTAACCTGAACCGCTGTGAAGCCACCA 120
 DB 661 GTCTATCTGCACGCGCGGACGCGCTTAACCTGAACCGCTGTGAAGCCACCA 720
 OY 121 CAGTTTACAGTCTGAGCGCGGGAACGTCATCGCAGCCAGTTTCCGGAAGCCCAACGC 180
 DB 721 CAGTTTACAGTCTGAGCGCGGGAACGTCATCGCAGCCAGTTTCCGGAAGCCCAACGC 780
 OY 181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTAAGTCAACCGGAT 240
 DB 781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTAAGTCAACCGGAT 840
 OY 241 GATTAACCTGCTGAGGAGCGCGGACGACCTGAATCCATCAACATGGTGGCAATGC 300
 DB 841 GATTAACCTGCTGAGGAGCGCGGACGACCTGAATCCATCAACATGGTGGCAATGC 900
 OY 301 TATCGCGTCCGCTGCAACCGGCGATGAGATTATTTGTCAGCGTGCAGAACACAC 360
 DB 901 TATCGCGTCCGCTGCAACCGGCGATGAGATTATTTGTCAGCGTGCAGAACACAC 960
 OY 361 GCCAACCTGCTCCCTGCGTGAATGGTGCACCAACAACTGGAGCCAAAGTGTGAATTG 420
 DB 961 GCCAACCTGCTCCCTGCGTGAATGGTGCACCAACAACTGGAGCCAAAGTGTGAATTG 1020
 OY 421 CCGCTTAATGCGAGCGACGCGCGGATGATTTGTGGCAGAACTGATTACTCCCGT 480
 DB 1021 CCGCTTAATGCGAGCGACGCGCGGATGATTTGTGGCAGAACTGATTACTCCCGT 1080
 OY 481 AGTCGATTCGCGTGGGTGAGATGTCGAACGTTACTGCGGTTGCCCGATCTGGC 540
 DB 1081 AGTCGATTCGCGTGGGTGAGATGTCGAACGTTACTGCGGTTGCCCGATCTGGC 1140
 OY 541 CGACGATTAACCTTCTCATTCACCGCGGATGCTGATGCTGCTCAAGCG 600
 DB 1141 CGACGATTAACCTTCTCATTCACCGCGGATGCTGATGCTGCTCAAGCG 1200
 OY 601 GCATGATTTCCCGGCGGATGTCAGCACTGATATTAATTTCTATGCTTTTCAGGT 660
 DB 1201 GCATGATTTCCCGGCGGATGTCAGCACTGATATTAATTTCTATGCTTTTCAGGT 1260
 OY 661 CACAACTGATGCGCGGACGATGCGGCTGATGATGTAATGCAACTGCTGAG 720
 DB 1261 CACAACTGATGCGCGGACGATGCGGCTGATGATGTAATGCAACTGCTGAG 1320
 OY 721 GCGATGCGCGGCTGCGTGGCGGCGCAAAATGCTTCAAGAGTGAATTTGACGCGTTC 780
 DB 1321 GCGATGCGCGGCTGCGTGGCGGCGCAAAATGCTTCAAGAGTGAATTTGACGCGTTC 1380

OY 781 AGCACTAATCTGCGCGGTGGAAGTGAAGCTGGAAGCCCAATGTCGTGCATTA 840
 DB 1381 AGCACTAATCTGCGCGGTGGAAGTGAAGCTGGAAGCCCAATGTCGTGCATTA 1440
 OY 841 GGATTAAGCGCGCGCTGCAATGCTGCGAGATTAAGATATCAACAGCCCAAGCTG 900
 DB 1441 GGATTAAGCGCGCGCTGCAATGCTGCGAGATTAAGATATCAACAGCCCAAGCTG 1500
 OY 901 AGCGTAGCTTGAACAGGCTGGCGGAGATGCGCTGGGGAACGTCGCGCTTCTGTTCA 960
 DB 1501 AGCGTAGCTTGAACAGGCTGGCGGAGATGCGCTGGGGAACGTCGCGCTTCTGTTCA 1560
 OY 961 TTCGCTGCGAGATTCACGCTGCGCTTGTGATTTGCTGCGCTTATACATACGAT 1020
 DB 1561 TTCGCTGCGAGATTCACGCTGCGCTTGTGATTTGCTGCGCTTATACATACGAT 1620
 OY 1021 ATGTCACGCTGCTGCGGAGTACGATATGCTGCGCGCGCGGACGATTCGCTCAG 1080
 DB 1621 ATGTCACGCTGCTGCGGAGTACGATATGCTGCGCGCGCGGACGATTCGCTCAG 1680
 OY 1081 CCGCTACTGCGAGATTAAGCGCTAACCGGACACGCTGCGCGCTTTGGCGCATTAAT 1140
 DB 1681 CCGCTACTGCGAGATTAAGCGCTAACCGGACACGCTGCGCGCTTTGGCGCATTAAT 1740
 OY 1141 ACAAGAGTGAATGATGCGCTGCTGTAATGCGCTTGACGCGCGCTGGAATTAATG 1200
 DB 1741 ACAAGAGTGAATGATGCGCTGCTGTAATGCGCTTGACGCGCGCTGGAATTAATG 1800
 OY 1201 GATTAA 1206
 DB 1801 GATTAA 1806

RESULT 5
 AA209792
 ID AA209792 standard; DNA; 4975 BP.
 XX
 AC AA209792;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Plasmid pHS1 mek/ biosI fusion protein DNA.
 XX
 KM SAM: S-adenosyl methionine synthase; biosynthesis; biotin; biosI; bios2;
 KW biosI; cofactor; decarboxylation; Vitamin H; mek; ss.
 XX
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT CDS 530..1684 /tag= a
 FT CDS 1782..2987 /product= "mek"
 FT CDS /tag= b
 FT /product= "biosI"
 XX
 PN DE19806872-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 19-FEB-1998; 98DE-1006872.
 XX
 PR 19-FEB-1998; 98DE-1006872.
 XX
 PA (BAD1) BASF AG.
 XX
 PI Schroeder H.;
 XX
 DR WP1: 1999-48095/41.
 XX
 DR P-PDB: AA133269, AA133270.
 XX
 PT Production of biotin by expressing S-adenosyl-methionine synthase and
 PT second biotin synthesis gene in host cells .

XX Example 3: Page 36-41; 48pp; German.
 XX This invention describes a novel method for the preparation of biotin
 CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
 CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
 CC sequence (1), and (b) at least one of the other biotin biosyntheses
 CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
 CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
 CC most animals and some microorganisms. Expression of biotin plus bios1,
 CC bios2 or bios3 significantly increases productivity of biotin
 CC biosynthesis, particularly by at least 3 times. This sequence encodes
 CC the meK/bios1 fusion protein found in plasmid pBI1 which is used in the
 CC method of the invention.
 XX
 XX Sequence 4975 BP; 1198 A; 1390 C; 1329 G; 1098 T; 0 other:
 SO
 Query Match 100.0%; Score 1206; DB 20; Length 4975;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAACGTTTTTAATCCCGCAGATTTCGCCGCCAGTTTCCCGCAGTACAGAGATCGGAGC 60
 DB 1782 ATGAACGTTTTTAATCCCGCAGATTTCGCCGCCAGTTTCCCGCAGTACAGAGATCGGAGC 1841
 OY 61 GTCATCTCGACAGCGCGCGCGCTTAACCTGAAGCGGTGTAAGCCACCCAA 120
 DB 1842 CTCTATCTCGACAGCGCGCGCGCTTAACCTGAAGCGGTGTAAGCCACCCAA 1901
 OY 121 CAGTTTACAGTCTGAGCGCGGAAACGTCATCGACAGCAGTTTGGCCAAAGCCCAACGC 180
 DB 1902 CAGTTTACAGTCTGAGCGCGCGGAAACGTCATCGACAGCAGTTTGGCCAAAGCCCAACGC 1961
 OY 181 CTGACCGCGCGCTTATGAAGCTGCACAGAGAAAGTGCGGCAGTACTAATGACCGGAT 240
 DB 1962 CTGACCGCGCGCTTATGAAGCTGCACAGAGAAAGTGCGGCAGTACTAATGACCGGAT 2021
 OY 241 GATAAAGTATGCTGTGAGCGGCGGACCACTGAATCATCAACATGATGGGCAACATGC 300
 DB 2022 GATAAAGTATGCTGTGAGCGGCGGACCACTGAATCATCAACATGATGGGCAACATGC 2081
 OY 301 TATGCGCGCTCGCGCTGTCAACCGGCGCATGAGATTATGTCAGCGTGGCAGAACCCAC 360
 DB 2082 TATGCGCGCTCGCGCTGTCAACCGGCGCATGAGATTATGTCAGCGTGGCAGAACCCAC 2141
 OY 361 GCCAACCTGTCGCCCTGCTGATGTGCGCCCAACAACCTGAGCCCAAGTGTGAATTCG 420
 DB 2142 GCCAACCTGTCGCCCTGCTGATGTGCGCCCAACAACCTGAGCCCAAGTGTGAATTCG 2201
 OY 421 CCGCTTATGGCGACGCACTGCGGATGTCGATTGTTGCCAGAACTGATTAATCCCGCT 480
 DB 2202 CCGCTTATGGCGACGCACTGCGGATGTCGATTGTTGCCAGAACTGATTAATCCCGCT 2261
 OY 481 AGTCGATTTCTGCGCTTGGGTGAGATGTCGAACGTTTACTGCGCGGTTGCCGATTCGCG 540
 DB 2262 AGTCGATTTCTGCGCTTGGGTGAGATGTCGAACGTTTACTGCGCGGTTGCCGATTCGCG 2321
 OY 541 CGAGCGATTACCTTGTCTCATTCACCGCGGATGTCGATGTTGATGTCGTCAGGCG 600
 DB 2322 CGAGCGATTACCTTGTCTCATTCACCGCGGATGTCGATGTTGATGTCGTCAGGCG 2381
 OY 601 GCAGTGCATTTCCCGCGGATGTTGCAAGTATGATTTCTATGCTTTTTCAGCT 660
 DB 2382 GCAGTGCATTTCCCGCGGATGTTGCAAGTATGATTTCTATGCTTTTTCAGCT 2441
 OY 661 CACAAAGTATGCGCCGACAGATTCGCGCTGTGTATGTAATTCGAACCTGTCGAG 720
 DB 2442 CACAAAGTATGCGCCGACAGATTCGCGCTGTGTATGTAATTCGAACCTGTCGAG 2501
 OY 721 GCGATGTGCGCTGCTGCGCGCGGCAAAATGCTTACGCAAGTGAATTTTGACGCGTTC 780
 DB 2502 GCGATGTGCGCTGCTGCGCGCGGCAAAATGCTTACGCAAGTGAATTTTGACGCGTTC 2561

OY 781 ACGACTCAATCTGGCGCGTGGAACTGGAAGCTGGAACGCCAAATGTCCTGTCTCAT 840
 DB 2562 ACGACTCAATCTGGCGCGTGGAACTGGAAGCTGGAACGCCAAATGTCCTGTCTCAT 2621
 OY 841 GGATTAACCGCGCGCTGGAATGCTGCGACATTACATATCAACACAGCCGAAAGCTGC 900
 DB 2622 GGATTAACCGCGCGCTGGAATGCTGCGACATTACATATCAACACAGCCGAAAGCTGC 2681
 OY 901 AGCGCTACCTTACGAACCGCTGCGGAAGATCCGTCGCGGAACGTCGCCGCTTTCCTCA 960
 DB 2682 AGCGCTACCTTACGAACCGCTGCGGAAGATCCGTCGCGGAACGTCGCCGCTTTCCTCA 2741
 OY 961 TTCCGCTCCGACGAAATTCACGCTGCTGCGCTTATGTTGTCGCTCATCAATACCAT 1020
 DB 2742 TTCCGCTCCGACGAAATTCACGCTGCTGCGCTTATGTTGTCGCTCATCAATACCAT 2801
 OY 1021 ATGTGACGCTGCTGCGCGAATACGCTTTCGCTGCGCGCGGACAGATTCGCTGAC 1080
 DB 2802 ATGTGACGCTGCTGCGCGAATACGCTTTCGCTGCGCGCGGACAGATTCGCTGAC 2861
 OY 1081 CCGCTACTGCGCAATTAAGCGCTAACCGGACACTGCGCGCTTTTGGCCCATATAT 1140
 DB 2862 CCGCTACTGCGCAATTAAGCGCTAACCGGACACTGCGCGCTTTTGGCCCATATAT 2921
 OY 1141 ACAAGAGTGAATGATGATGCGCTGCTGTAATCCGTTGACCGCGGCTGGAATTAATG 1200
 DB 2922 ACAAGAGTGAATGATGATGCGCTGCTGTAATCCGTTGACCGCGGCTGGAATTAATG 2981
 OY 1201 GATTAA 1206
 DB 2982 GATTAA 2987

RESULT 6
 AAX02810
 ID AAX02810 standard; DNA: 1216 BP.
 XX
 AC AAX02810;
 XX
 DT 14-MAY-1999 (first entry)
 XX
 DE E. coli biotin ORF401 DNA.
 XX
 KW Biotin: ORF401; synthesis: dechlobiotin; yield: increase; industry:
 KW fermentation; ss.
 XX
 OS Escherichia coli.
 XX
 XX
 FT Key Location/Qualifiers
 FT CDS 12..1216
 FT /"tag" a
 FT /product= "ORF401"
 FT
 XX
 PN DE19731274-A1.
 XX
 PD 28-JAN-1999.
 XX
 XX 22-JUL-1997; 97DE-1031274.
 PF
 XX 22-JUL-1997; 97DE-1031274.
 PR
 XX (BAD1) BASF AG.
 PA
 XX Hauer B, Schroeder H;
 PI
 XX
 DR WPI: 1999-107030/10.
 DR P-PSDB: AAM92934.
 XX
 PT Improved synthesis of biotin by expressing the bios1 or bios2
 PT sequence in biotin-producing cells - and related gene constructs,
 PT provides increased conversion of dechlobiotin in eukaryotic or
 PT prokaryotic hosts
 XX

PS Claim 1: Page 12-14; 48pp; German.
 XX This invention describes a method for the synthesis of biotin in
 CC Escherichia coli. This method involves the expression of a biotin gene
 CC or its functional variants, analogues and derivatives, in a prokaryote
 CC or eukaryote that is able to produce biotin. The cells are grown and
 CC the biotin produced either used directly, after separation of the
 CC biomass, or after purification. Constructs containing this nucleic acid
 CC or protein or their variants etc., can be coupled to one or more
 CC regulators for increasing gene and/or protein expression, and/or having
 CC its natural regulators 'switched off'. Expression of this biotin protein
 CC leads to increased conversion, by at least 3-fold, of dehydrobiotin to
 CC biotin, thus increasing yield and making possible an industrially useful
 CC fermentative method for biotin production.
 XX
 XX Sequence 1216 BP; 258 A; 318 G; 363 C; 277 T; 0 other:

Query Match 99.9%; Score 1205; DB 20; Length 1216;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACGCTTTTAAATCCCGGAGTTTCCGCCCACTTCCGCACTACAGATGCGGC 60
 DB 12 ATGAAAGTTTAAATCCCGGAGTTTCCGCCCACTTCCGCACTACAGATGCGGC 71
 OY 61 GCTATCTGACAGCGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 120
 DB 72 GCTATCTGACAGCGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 131
 OY 121 CAGTTTACAGTGTAGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 180
 DB 132 CAGTTTACAGTGTAGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 191
 OY 181 CTGACCGCGGTTTAAAGCTGACAGAGAAAGTGGCCAACTTACTGATGACCGGAT 240
 DB 192 CTGACCGCGGTTTAAAGCTGACAGAGAAAGTGGCCAACTTACTGATGACCGGAT 251
 OY 241 GATAAAGTATGCTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 300
 DB 252 GATAAAGTATGCTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 311
 OY 301 TATGCGCGTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 360
 DB 312 TATGCGCGTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 371
 OY 361 GCCAAGCTGTGCTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 420
 DB 372 GCCAAGCTGTGCTGTGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 431
 OY 421 CGGCTTAAAGCGGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 480
 DB 432 CGGCTTAAAGCGGAGCGCGGAGTTTAAAGCGGTGTTGAAGCCACCA 491
 OY 481 AGTGGATTTCTGCGGTGAGATGTAAGCTTACTGCGGTTGCCGAGTCTGGCG 540
 DB 492 AGTGGATTTCTGCGGTGAGATGTAAGCTTACTGCGGTTGCCGAGTCTGGCG 551
 OY 541 CGAGCGATTAATCTTCTCTATGAGCGGAGTGTGATGTTGCTCAAGGG 600
 DB 552 CGAGCGATTAATCTTCTCTATGAGCGGAGTGTGATGTTGCTCAAGGG 611
 OY 601 GCAGTGAATTTCCCGGAGTTTAAAGCGGTGTTGAAGCCACCA 660
 DB 612 GCAGTGAATTTCCCGGAGTTTAAAGCGGTGTTGAAGCCACCA 671
 OY 661 CACAAGCTGATGCGCGGAGTATGCGGCTGTATGTAATCAGAACTGTGAG 720
 DB 672 CACAAGCTGATGCGCGGAGTATGCGGCTGTATGTAATCAGAACTGTGAG 731
 OY 721 GCAGTGTGCGGCTGTGCGGCGGAGTATGTAATCAGAACTGTGAG 780
 DB 732 GCAGTGTGCGGCTGTGCGGCGGAGTATGTAATCAGAACTGTGAG 791

OY 781 ACGACTAATCTGCGCGGAGTAACTGAGAGCTGGAACCCAAATGTCGTGTATTA 840
 DB 792 ACGACTAATCTGCGCGGAGTAACTGAGAGCTGGAACCCAAATGTCGTGTATTA 851
 OY 841 GGATTAAGCGCGGCTGGAATGCTGCGAGATATGATATGACAGCGCCGAAGCTGG 900
 DB 852 GGATTAAGCGCGGCTGGAATGCTGCGAGATATGATATGACAGCGCCGAAGCTGG 911
 OY 901 ACCCGTACCTTGAACAGCTGCGGAGATGCGGCGGAGTAAAGCTGCGGCTTCTTCA 960
 DB 912 ACCCGTACCTTGAACAGCTGCGGAGATGCGGCGGAGTAAAGCTGCGGCTTCTTCA 971
 OY 961 TTCGCTGCGGAGATGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1020
 DB 972 TTCGCTGCGGAGATGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1031
 OY 1021 ATGCTACCGCTGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1080
 DB 1032 ATGCTACCGCTGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1091
 OY 1081 CCGCTACCTGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1140
 DB 1092 CCGCTACCTGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1151
 OY 1141 ACAAGAGTATGATGATGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1200
 DB 1152 ACAAGAGTATGATGATGCGGAGTAAAGCTGCGGCGGAGTAAAGCTGCGGCTTCTTCA 1211
 OY 1201 GATTA 1205
 DB 1212 GATTA 1216

RESULT 7
 ABQ48374
 ID ABQ48374 standard; DNA; 524 BP.

ABQ48374;
 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34965.
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guehl D;

DR WPI: 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA

PS Claim 12: 56pp + Sequence Listing: 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (1) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

S0 Sequence 524 BP; 114 A; 44 C; 152 G; 214 T; 0 other;

Query Match	30.68;	Score 369;	DB 24;	Length 524;
PostgreSQL Client	01.08.2024	08:00	01.08.2024	08:00

Best Local Similarity 81.8%; Pred. No. 3.1e-106

Matches 426; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY	119	AACGTTTTACAGTCGAGGCCCGGAAGCCCTCCAGCAGCATTTTCCGAACCCAC	178
Db	4	AATAGTTTTATAGTTTGAGCGCTCGGAACGTTTATCGTAGTAACTTTCTCGAATTAA	63
OY	179	GCCGTACCCCGCGTTATGAACTGCACAGAGAAAGTGGCGCAATTACTAATGCACGG	238
Db	64	GTTTGATTCGCGCGTTATGCAAGTTGTACAGAGAAAGTGGCGCAATTACTAATGCACGG	123
OY	239	ATGATTAACACTATGCTGAGACGGCGGACCACTGAAATCATCAATGATGGTGCACAT	299
Db	124	ATGATTAATATTATGCTTTTGGACGGCGCGTATATTGAATTTATTAATATGTTGATTAAT	183
OY	299	GCTATGGCGCGTCCGCGTCCACACGGGCGATGAGATTATTGTCAGCGTGGCAGACACC	358
Db	184	GTTATGCGCGTTCCCGTTTGAAATCGGGCGATGAGATTATTGTTAGCGTGTAGAAATTT	243
OY	359	ACGCCAACCTGTCGCCGCTGATGTCGCCCAACAACGGAGCGCAAAAGTGTGAAT	418
Db	244	ACGTTAATTTTCGTTTTTTTGGTGAATGTCGTTTAAATAATTTGGAGTTAAAGTGTGAAT	303
OY	419	TGCGGCTTAATGCGACAGCAGTCCGGATGTGATTTGTGCCGAACGTATTACTCCC	478
Db	304	TGTGCTTAATGCGTAGAGATTTGCGAGATGTCGATTTGTGTAATTTGATTTTTC	363
OY	479	GTAATCCGATTCGCGGTTGGGTGAGATTCGAATTCGAACGTTACTGGCGGTTCCCGGATCTGG	538
Db	364	GTAATCCGATTTGGCGTTGGGTGAGATTCGAACGTTATGGGCGGTTGTCGGATTTGG	423
OY	539	CGCGAGCATATCCCTTTGCTATTACACCGGAGTGTGTGATGTTGATGCTCTCAG	598
Db	424	CGCGAGCATATTTTGTATTATTAGTCGGAGTGTGTGATGTTGATGCTTTAGG	483
OY	599	GGCGAGTCATTTCCCGCGGATGTTTCAAGCAACTGAGATTT	659
Db	484	CGGTAGTGTATTTTTTTCGGGAGCTTTAGTAATTTGCAATTT	524

RESULT 8
AB040375 /5

ID ABQ48375 standard; DNA; 524 BP.

AC ABQ48375;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34966

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

OS Homo sapiens.
xy

PN WO200218632-A2
XY

PD 07-MAR-2002
yy

PF 01-SEP-2001; 2001WO-EP10074.
VY

PR	01-SEP-2000; 2000DE-1043826.
BP	05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPICENOMTOS AC

XX	Olak A	Pionobrock C
PT		

XX WPT: 3003-371838/40
DB

XX
XX
DE

Determine if you + have access

PT for diagnosis and prognosis, comprises selective hybridization of

PS Claim 12; 56pp + Sequence Listing; 56pp; German

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABO13110-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 524 BP; 214 A; 152 C; 44 G; 114 T; 0 other;

Query Match	Score	DB	Length
30.68	369	24	524

Best Local Similarity 81.8%; Pred. NO. 3.1e-106;
Matches 436; Conservative 0; Mismatches 95; Indels 0;

Matches	426;	Conservative	0;	Mismatches	95;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	-----	--------	----	------	----

Oy	119	AACAGTTTACAGTCGTAGGCCCCGGAAACGTCATCTCAGCAGCAATATTTGGCGGAAGCCCCAAC	178
Db	521	AATAGTTTATAGTTTAGAGCTCGGAACCTTTATCGAGTTAATTTGTGTGCAAGTTTAA	462
Oy	179	GCGTCACCCGCGGTTATGAACTCAGACAGCAAGAAAGTGGCGCAATTACTGAATGCACCCG	238
Db	461	GTTTGATATGCGCGTTATGAACTTTATGACAGCAAGAAAGTGGCGCAATTATGAAATGTATACGG	402
Oy	239	ATGATAAACATATGCTGTGAGCGCGGCAACACTGCAATCCATCAACATGGTGGCAAT	298
Db	401	ATGATTAATTTATGCTTTGAGACCGCGCGTATATTGAAATTTATTAATATATGTGTGTATAAT	342
Oy	299	GCTATGCGCGTCCGCGTCGACACCGGGCGATGATTTATTTGACCGTGGCAGACACC	352
Db	341	GTTATGCGCGTTCGCGTTTGTAATCGGGCGATGATGATTTATTTGACGCTGGTAAATATT	282
Oy	359	ACGCCAACCTCGTCCCGCTGATATGTCGCCCAACAACGAGGAGCCAAAGTGGTGAAT	418
Db	281	ACGTTAATTTGCTTTTGGTGTGATAGTGCCTTTATATAATTTGAGGTTTAAAGTGGTGAAT	222

QY 419 TCCGCTTAATGCGACGACGATCCGATGCTATTGTTGCCAGAACTACTACCC 478
 DB 221 TCGCTTTATGCGTACGATGCTGATGCTATTGTTGATGATGATTTTC 162
 QY 479 GTATGCGATTTGCGGCTGCGTACAGATGCAACGTTACTGCGGTTCCCGATCTCG 538
 DB 161 GTAGTCGATTTGGGCTGGGTAGATGCAACGTTATGGCGGTGTTGCGATTTGG 102
 QY 539 CCGGACGATTTACTTTGCTGATTCAGCCGGATGGTGGTGGTGGTGGTGGTGGTGG 598
 DB 101 CCGGACGATTTATTTGTTATTTAGTCGGATGGTGGTGGTGGTGGTGGTGGTGG 42
 QY 599 GGGCAGTGCATTTCCCGGATGTTGAGAACTGATATT 639
 DB 41 GGGTACTGATTTTTCGCGGATGTTAGTAATGATATT 1

 RESULT 9
 ABO48376/c
 ID ABO48376 standard; DNA: 524 BP.
 AC ABO48376;
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34967.
 XX
 KM Human: cytosine methylation; 5'-CpG-3': uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 PN WO200218632-A2.
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001: 2001WO-EP10074.
 XX
 PR 01-SEP-2000: 2000DE-1043826.
 PR 05-SEP-2000: 2000DE-1044543.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX
 PS Claim 12: 56bp + sequence listing; 56bp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP-s); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.

CC ABO13410-AB054121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 SQ Sequence 524 BP; 116 A; 44 C; 142 G; 222 T; 0 other:
 Query Match 28.9%; Score 348; DB 24; Length 524;
 Best local similarity 79.0%; Pred. No. 1.4e-99;
 Matches 414; Conservative 0; Mismatches 110; Indels 0; Gaps 0:
 QY 116 CCCAAGATTTTACATCTGAGCGCGGAAACGTCATGCGCACTGATGCGCAAC 175
 DB 524 CCCAAGATTTTACATCTGAGCGCGGAAACGTCATGCGCACTGATGCGCAAC 465
 QY 176 AAGCGCTGACCGGCGGTTTGAAGCTGACGAGAGAAAGTGGCAATTAAGTACGAC 235
 DB 464 AAGCGCTGACCGGCGGTTTGAAGCTGACGAGAGAAAGTGGCAATTAAGTACGAC 405
 QY 236 CGGATGATTAACATATCTGCTGACGCGCGGCACTGATGCGCAACATGATGCGCAC 295
 DB 404 CGAATTAACATATCTGCTGACGCGCGGCACTGATGCGCAACATGATGCGCAC 345
 QY 296 AATGCTATGCGCGCTCCGCTCTGCAACGCGGCGATGATTTATGTCAGCGTGGCAGAC 355
 DB 344 AATGCTATGCGCGCTCCGCTCTGCAACGCGGCGATGATTTATGTCAGCGTGGCAGAC 285
 QY 356 ACCAGCGCAACCTGCGTCCCTGCTGATGCTGCGCCCAACAACTGGAGCGCAAGTGTGA 415
 DB 284 ACCAGCGCAACCTGCGTCCCTGCTGATGCTGCGCCCAACAACTGGAGCGCAAGTGTGA 225
 QY 416 AATTGCGGCTTAATGCGACGACGCTCCGATGCTGATTTGTTGCCAGAACTGATTAC 475
 DB 224 AATTGCGGCTTAATGCGACGACGCTCCGATGCTGATTTGTTGCCAGAACTGATTAC 165
 QY 476 CCCGTAAGTGGATTCGCGCTGGGTCAGATGTCGAACGTTACTGCGCGTGGCCGATC 535
 DB 164 CCCGTAAGTGGATTCGCGCTGGGTCAGATGTCGAACGTTACTGCGCGTGGCCGATC 105
 QY 536 TGGCGGAGGATTTACTCTTTCGCTATTCAGCCGGATGGTGGTGGTGGTGGTGGTGG 595
 DB 104 TAAAGCGAAGATTTACTCTTTCGCTATTCAGCCGGATGGTGGTGGTGGTGGTGGTGG 45
 QY 596 AGGCGCAGTGCATTTCCCGCGGATGTTGAGCAACTGATATT 639
 DB 44 AAAAAACAATACATTTCCCGCGGATGTTGAGCAACTGATATT 1

 RESULT 10
 ABO48377
 ID ABO48377 standard; DNA: 524 BP.
 AC ABO48377;
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34968.
 XX
 KM Human: cytosine methylation; 5'-CpG-3': uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 PN WO200218632-A2.
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001: 2001WO-EP10074.
 XX
 PR 01-SEP-2000: 2000DE-1043826.
 PR 05-SEP-2000: 2000DE-1044543.
 XX

PA	(EPIC-) EPICENOMICS AG.
Pt	Olek A., Piepenbrock C., Berlin K., Gueutig D;
DR	WPI: 2002-371829/40.
XX	Determining the degree of cytosine methylation in genomic DNA, useful
PT	for diagnosis and prognosis, comprises selective hybridization of
PT	amplicons from chemically treated DNA -
XX	Claim 12: 56pp + Sequence Listing; 56pp; German.
PS	
CC	This invention describes a novel method for determining the degree of
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one
CC	member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC	and the degree of hybridisation to both classes is determined from the
CC	label on the amplicon. From the ratio of labels hybridised to the two
CC	classes of oligomers, the degree of methylation is calculated. The method
CC	is described as:
CC	CC (1) for diagnosis and/or prognosis of side effects of
CC	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC	of the central nervous, cardiovascular, gastrointestinal and respiratory
CC	systems etc., particularly by detecting mutations or single nucleotide
CC	polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
CC	types and for investigating cell differentiation. The method allows the
CC	methylation status of many C residues to be determined simultaneously.
CC	AB013107-AB014121 represent genomic DNA sequences used to illustrate the
CC	method for determining the degree of cytosine methylation described in
CC	the disclosure of the invention.
XX	
SQ	Sequence 524 BP: 222 A; 142 C; 44 G; 116 T; 0 other;
Query Match	28.9% Score 348; DB 24; Length 524;
Best Local Similarity	79.0%; Pred. No. 1.4e-99;
Matches 414; Conservative	0; Mismatches 110; Indels 0; Gaps 0;
OY	116 CCCAACACTTTTACAGTCGACGCCCGGAACGTGCATCGACGCCAAGTTGGCCGAAGCCC 175
DB	1 CCCAACACTTTTACAACTTAACGCCGGAAGCGTCATCGCAACCAATTACCGAAGCCC 60
OY	176 AACGCTGACCGCGCGGTATGACGTGCACGAGAAGAAAGTGCGCAATTTACTGAATGCAC 235
DB	61 AACCCCTAACCGCGCGGTATAAAGTACACGAAAAAATAAGCAATTACTAAAATACAC 120
OY	236 CGGATGNTAAAGTATGCTGTGACGCCGCGCACACAGTGAATCATCAACATGATGGCAC 295
DB	121 CGAATTAACCTAATACATATCTCTTAACAGCCCGACGACCACTAATATCCATCAACAATAATACAC 180
OY	296 AATGCTATGCGCGTCCGCGTGTGCAACCGCGGCGATGAGATTAATTTGACGCGGAGAAC 355
DB	181 AATTAATTAAGCGCGTCCGCGTGTGCAACCGGACGATTAATTAATTTGCAACGTAACCAAAAC 240
OY	356 ACACAGCCACACCTGTCGCTGTGCTGATGTGTCGCCCAACAACGTGAGCCCAAGTGTCGA 415
DB	241 ACCACGCCAACCTGCTCCCTCACTAACAATAAATTCGCCCAACAACCAAAAATAATAAA 300
OY	416 AATGCGCGCTTAATGCGGAGGACATCGCGCGGANTGTGATTTGTGGCAACAAGTATTAAC 475
DB	301 AATTACCGCTTAATTAAGCAACGACACTACCGAATATTCGATTTATTACCAAACCTAATTATC 360
OY	476 CCGCTAGTCGCAATTTGCGCTGTGCTGACAGATGTCAACGTTACTGCGGTTGCCGATC 535
DB	361 CCGCTAATCGAATTTACTAAGCTTAATCAATAATATGCAAGCTTAACGATTAACCCGAATC 420
OY	536 TGCGCGCAGCATTAACCTTTGCTCATTCACGCCGCGGATGGTGATGCTGTGATGGTGTCTC 595
DB	421 TAAACGCGCAAGTAACTTTACTACTCATCAACGGAATAATAATAATAATAATAATATAC 480
OY	596 AGGGGCGAGTGCATTTCCCCCGGATGTTCAACAACCTGATATT 639

Db	481	AAAAACAATACATTTCGCCCGGAGTATATCAACACTAAATTTT	524
		RESULT 11	
Xx	ID	ABO23332	
Xx	XX	ABO23332 standard; DNA; 917 BP.	
Xx	AC	ABO23332;	
Xx	DT	12-JUL-2002 (first entry)	
DE	XX	Oligonucleotide for detecting cytosine methylation SEQ ID NO 9923.	
XX	KW	Human; cytosine methylation: 5'-CpG-3'; uracil; cytosine; diagnosis;	
KW	KM	drug; side effect; cancer; central nervous system; cardiovascular;	
KW	KW	gastrointestinal; respiratory system; single nucleotide polymorphism;	
SNP; cell differentiation; ds.			
OS	XX	Homo sapiens.	
PN	XX	WO200218632-A2.	
PD	XX	07-MAR-2002.	
PF	XX	01-SEP-2001; 2001WO-EPI0074.	
PR	XX	01-SEP-2000; 2000DE-1043826.	
PR	XX	05-SEP-2000; 2000DE-1044543.	
PA	(EPIC-) EPIGENOMICS AG.		
PI	Olak A, Piepenbrock C, Berlin K, Guetig D;		
DR	WPI; 2002-371829/40.		
PT	Determining the degree of cytosine methylation in genomic DNA, useful		
PT	for diagnosis and prognosis, comprises selective hybridization of		
PT	amplicons from chemically treated DNA -		
XX	Claim 12: 56pp + Sequence Listing; 56pp; German.		
PS	This invention describes a novel method for determining the degree of		
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a		
CC	genomic sample of DNA. The sample is treated chemically to convert		
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic		
CC	DNA that contains the target C is amplified to form a labeled amplicon.		
CC	The amplicon is hybridised to two classes, each with at least one		
CC	member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers		
CC	and the degree of hybridisation to both classes is determined from the		
CC	label on the amplicon. From the ratio of labels hybridised to the two		
CC	classes of oligomers, the degree of methylation is calculated. The method		
CC	is used: (i) for diagnosis and/or prognosis of side effects of		
CC	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders		
CC	of the central nervous, cardiovascular, gastrointestinal and respiratory		
CC	systems etc., particularly by detecting mutations or single nucleotide		
CC	polymorphisms (SNPs); and (ii) for differentiation of cell or tissue		
CC	types and for investigating cell differentiation. The method allows the		
CC	methylation status of many C residues to be determined simultaneously.		
CC	AB013410-AB054121 represent genomic DNA sequences used to illustrate the		
CC	method for determining the degree of cytosine methylation described in		
CC	the disclosure of the invention.		
SO	Sequence 917 BP; 201 A; 85 C; 250 G; 381 T; 0 other;		
	Query Match	23.8%; Score 287.6; DB 24; Length 917;	
	Best Local Similarity	79.1%; Pred. No. 2.4e-80;	
	Matches 478; Conservative	0; Mismatches 114; Indels 12; Gaps 11.	
OY	9	TTTTAATCCCGCGCATTTCCGGCCCACTTCCCACATACAGATGCGGCCTATCT	68
Db	71	TTTATAATTTCGCGTAGTTTCGCGTTACTTTTTCGTATTATAGANTCGCGCGCT-TATT	129
Yy	69	CGACAGCGCGCGAGCCCGCTTAACCTGAAGCCGTGCTTGAGGCCACCACCACTTTTA	128

```

Db 130 CGATTACGTCGGATCGGATCGTTAAATTTGAAGTCGTGAA-TTATTTAATAGTTTA 188
Oy 129 CAGCTGACCGCCGAAACCTCATGCGACCCAGTTGCCGAAGCCCAACGCTGACCC 188
Db 189 TAGTTGACCGTCGAAAGCTTATTCGTAGTTA-TTTGTCGAAGTTTAACTGATCC 247
Oy 189 GCGTTGAAAGTCGACGAGAAAGTGGCGCAATAGTCAATGACGCGATGATAAAG 248
Db 248 GCGTTGAAAGTCGACGAGAG-AACTGCGCTAATTATTTGAATGATCGATGATAAAT 306
Oy 249 TATCGTGTGACGCGCGACCACTGAATCATCAACATGTGGCACAATGCTATGCGCG 308
Db 307 TATCGTTTGACG-CGCGTATTTATTTAATTTATATGCTGATATGTTATGCGCG 365
Oy 309 TCCGCGTGTGCAACCGCGCGATGATTTATTTGTCACCGTGGCGAAGACCCCAACT 368
Db 366 T-TTCGCTTTGTAATCGCGCATGACATTTATTTACGCTGTAATTTACTAATTC 424
Oy 369 CGTCCCGCTGCTGATGCTGGCCCAACAACCTGAGCAAGTGAATTCGCGCTTA 428
Db 425 GTT-TTGGTGTATGCTGCTTTAATTAATTTGAAGTTAG-CGAATTTGCTTTAA 481
Oy 429 TCCGACGACGCTGCGGATGCTGATTTGTTGCCAGAACTGATCTCCCGTATGCGAT 488
Db 482 TCGGTACGATGCTGCGATGCTGATTTGTTGA-TAATGATATTTTTCGTAGTCGAT 540
Oy 489 TCTGCGCTTGGCTCAGATGTCGAACTTACTGCGGCTTCCCGGATCTGCGGACGAT 548
Db 541 TTTGCGCTTGGCTTACATGTCGAA-GTTATGCGGCTTTCGATTTTGGCGGACGAT 599
Oy 549 TACTTTCGCTTACAGCGCGGATGATGATGATGATGATGATGATGATGATGATGATG 608
Db 600 TATTTTGTATTAT-AGTCGCGATGCTGATGATGATGATGATGATGATGATGATG 658
Oy 609 TTTC 612
Db 659 TTTC 662

RESULT 12
ABO23333/c
ID ABO23333 standard; DNA: 917 BP.
XX
XX ABO23333:
XX
XX 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9924.
XX
XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
XX MO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001: 2001MO-EP10074.
XX
XX 01-SEP-2000: 2000DE-1043826.
XX
XX 05-SEP-2000: 2000DE-104543.
XX
XX (EPIC-) EPIDENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guelzig D:
XX
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT

```

```

PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA.
XX
XX Claim 12: 56pp + Sequence Listing: 56pp: German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridized to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridization to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridized to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (1) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (11) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 917 BP: 381 A; 250 C; 85 G; 201 T; 0 other:
XX
XX Query Match 23.8%; Score 287.6; DB 24: Length 917:
XX Best Local Similarity 79.1%; Pred. No. 2,4e-80:
XX Matches 478; Conservative 0; Mismatches 114; Indels 12; Gaps 11:
Oy 9 TTTTATCCCGGCGCATTTCCGCGCCGCTTAACCTGAAGCGGTGAAGCCACCCACAGTTTA 68
Db 847 TTTTATTTTCGCTAGTTTCGCTTTAGTTTTCGATTTATAGATGCGGCGCT-TATTT 789
Oy 69 CGACAGCGCGCGACCGCGCTTAACCTGAAGCGGTGAAGCCACCCACAGTTTA 128
Db 788 CGATAGCGTCGCGATGCGCTTAATTTGAAGTCGTGTTGA-TTATTTAATAGTTTAA 720
Oy 129 CAGCTGAGCGCGGAAACCTCCATGCGACGCTTCCCGAAGCCCAACGCTGACCG 188
Db 729 TACTTTGACGCTGCGAAACCTTTATCTGATTTA-TTTGTCGAAGTTAAGTTGATCCG 671
Oy 189 GCGTTATGAAGTCGACGAGAGAAAGCGGCAATTAAGTAATGACCGGATGATAAAG 248
Db 670 GCGTTATGAAGTCGACGAGAG-AACTGCGCTAATTGAAGTAATGCGATATAAAT 612
Oy 249 TATCGTGTGACGCGCGACCACTGAATCATCAACATGCTGGCAACATGCTATGCGCG 308
Db 611 TATCGTTTGACG-CGCGTATTTTATTAATTTATATGCTGATATATGCTATGCGCG 553
Oy 309 TCCGCGTGTGCAACCGCGCGATGATTTATTTGTCACGCTGGCGAAGACCCCAACT 368
Db 552 T-TTCGCTTTGTAATCGCGCATGAGATTAATTTAGCGCTGATATATTAATTAATTC 494
Oy 369 CGTCCCGCTGATGCTGCCCAACAACCTGAGCAAGTGTGAATTTGCCGCTTA 428
Db 493 GTT-TTGGTGTATGCTGCTTTAATTAATTTGAAGTTAG-CGAATTTGCTTTAA 427
Oy 429 TCCGACGACGCGCGGATGCTGATTTGTTGCCAGAACTGATTAATCTCCCGTATGCGAT 488
Db 436 TCGGTACGATGCTGCGATGCTGATTTGTTGA-TAATGATATTTTTCGTAGTCGAT 378
Oy 489 TCTGCGCTTGGCTCAGATGTCGAACTTATGCGGCTTCCCGGATCTGCGGACGAT 548
Db 377 TTTGCGCTTGGCTTACATGTCGAA-GTTATGCGGCTTGTTCGATTTTGGCGGACGAT 319
Oy 549 TACTTTCGCTTACAGCGCGGATGCTGATGATGCTGATGCTGCTCAGCGGACGATGCA 608
Db 318 TATTTTGTATTAT-AGTCGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 260
Oy 609 TTTC 612
PT

```

DB 259 TTTC 256

RESULT 13

AB023330/C
ID AB023330 standard; DNA: 917 BP.

XX AB023330:

DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 9921.

XX Human: cytosine methylation, 5'-CpG-3'; uracil: cytosine; diagnosis;
KW drug; side effect: cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guelzig D;

XX WPI: 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
PS
PS Claim 12: 56pp + Sequence Listing: 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridized to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridization to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridized to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNPs); and (11) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB013410-AB034121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 917 BP: 202 A; 85 C; 264 G; 366 T; 0 other:

XX Query Match 23.0%; Score 277; DB 24; Length 917;

XX Best Local Similarity 77.6%; Pred. No. 5.5e-77;

DB 852 AACGTTTATCCCGCATTTCCGCCCATTTCCCGCATTTCAAAATACGACG-C 794

QY 64 TATCTGACAGCCCGCGCATTTAAACCTGAAGCCGTGTGAGGCCACCCACAG 123
DB 793 TATCTGACAGCCCGCGCATTTAAACCTGAAGCCGTGTGAGGCCACCCACAG 735
QY 124 TTTTACAGCTGAGCGCGGAAAGCTCCATGCGAGCCAGTTGGCGAAGCCCAACGCTG 183
DB 734 TTTTACAGCTGAGCGCGGAAAGCTCCATGCGAGCCAGTTGGCGAAGCCCAACGCTG 676
QY 184 ACCGCGCTTATGAGCTGACGAGAGAAAGCTGGCGAATTAAGTAATGACACCGATGAT 243
DB 675 ACCGCGCTTATGAGCTGACGAGAGAAAGCTGGCGAATTAAGTAATGACACCGATGAT 617
QY 244 AAAACTATCTGCTGAGCGCGCGGACACACATGATCATCAACATGCTGGCACAAATGCTAT 303
DB 616 AAAACTATCTGCTGAGCGCGCGGACACACATGATCATCAACATGCTGGCACAAATGCTAT 558
QY 304 GCGGCTGCGGCTGCGACACGCGGAGATGATTTATGTCAGCGTGGCGAGACACACGCGC 363
DB 557 ACGGCTGC-CGTCTTACACGCGGAGATTAATTTATATCAACGTAACAAACACGCGC- 500
QY 364 AACCTGCTCCCTGCTGATGCTGCGCCMACAACTGAGCGCAAGTGTGAATTTGCGC 423
DB 499 -AACCTGCTCCCTGCTGATGCTGCGCCMACAACTGAGCGCAAGTGTGAATTTGCGC 442
QY 424 CTTAATGCGGACGACGCGCGGATGCTGATTTGTCAGACTGATTAATCCCGCTAGT 483
DB 441 CTTAATGCGGACGACGCGCGGATGCTGATTTGTCAGACTGATTAATCCCGCTAGT 383
QY 484 CGGATTCGCGCTGGGCGATGCTGCAACCTTACTGCGGCTGCGCGGATTCGCGCGA 543
DB 382 CGAATTCGCGCTGGGCGATGCTGCAACCTTACTGCGGCTGCGCGGATTCGCGCGA 324
QY 544 GCGATTAACCTTGTCTCATTTACGCGCGGATGCTGATTTGTCAGACTGATTAATCCCGCTAGT 603
DB 323 ACGATTAACCTTGTCTCATTTACGCGCGGATGCTGATTTGTCAGACTGATTAATCCCGCTAGT 265
QY 604 GTGCATTTCCG 614
DB 264 ATACATTTCCG 254

RESULT 14

ID AB023331 standard; DNA: 917 BP.

XX AB023331:

DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 9922.

XX Human: cytosine methylation, 5'-CpG-3'; uracil: cytosine; diagnosis;
KW drug; side effect: cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guelzig D;

XX WPI: 2002-371829/40.


```
DB 196 ACCGAGAAATGAGAACGCTGCGCAAGCGGCATCGCTTTATTATATGCCGTTGCGCG 255
OY 244 AAAACTATGCTCTGAGCGCGCGGACCACTGAATCCATCAACATGCTGGCACATGCTAT 303
DB 256 GAAGAGCTGTGTTGCTCGCGGCGACGAGGAGGATCAATCTGTCGCCAATAGCTGG 315
OY 304 GCGCGTCCGCGTCTGCAACCGGGCATGAGATTATTGACGGTGGCAGAACCCAGCGCC 363
DB 316 GGCACACGACCACTGCGGCGCGGCGATACATCATCATGAGTGCAGTGCAGCACCCT 375
OY 364 AACCTGTCGCCCTGAGTGGTGGCCCAACAACACTGGAGCCCAAGTGTGAATTGGCG 423
DB 376 AACATGTTTCCCTGCGACATGCTTTGCGCAGCGCTTGGCGCAGACCTGCGTGTATCCG 435
OY 424 CTTAATGGCGACGACTGCGGATGCTGATTTGTTGCCAGACTGATTACTCCCGTAGT 483
DB 435 CTCATGCCGATGTGACTGTCAGTGCAGACGCTGCGCTGCTTGTGATGAGAAACT 495
OY 484 CGGATTCGGCGTGGGTCAGATGTCGAACCTTACTGGCGGTGGCCGATCTGGCGCA 543
DB 496 CGCGTGTGGCAATTACTGATGTCGCAACGCTTGGCAGACAGAAATCCACTGGCGGA 555
OY 544 GCGATTACCTTTGCTCATTCAGCGCGGATGGTGTGATGCTGATGCTCAGGGGCA 603
DB 556 ATGATCAGCGCTTGGCGCAGCAGATGCGCAAAAGTCTGTGTGATGCGCGCTCAGCGGTG 615
OY 604 GTGCATTTCCCGCGGATGTTTCAGCACTGATGATTTCTATGCTTTTTCAGGTCAC 663
DB 616 ATGCATCATCGCTGTGATGTTTCAGCGCGCTGATGCTGACTTTTACGTTTCCGCGCAT 675
OY 664 AAACCTATGCGCCGACAGTATCGCGCTGTATGTTGTAATCAGAACTGCTGAGGCG 723
DB 676 AAACGTATGCGCCGACGGAATTGGCATTTCTTATGTAAGAAGCCTTGTTCAGGAG 735
OY 724 ATGTGCGCTTGGCGCGCGGCAAAATGTTTCAGAAAGTAGTTT--TGACGCTTC 780
DB 736 ATGCGCGCTGTGGAAGGGCGGCTTCTATGATCGCCACCGCTGAGTGAAGGCACT 795
OY 781 ACGACTCAATCTGCGCGCTGGAACCTGGAAGCTGGAACGCCAAATGTGCTGTGCATA 840
DB 796 ACCTGACCAAAAGCACATGCGGCTTTGAAGCGGTACACCAATACCGGGGCACTCAT 855
OY 841 GGATTAAAGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACGAGCCGAAAGCTGG 900
DB 856 GGTCTTGGCGCGCGCTGAGATGTTTGGCGCTGGGGCTTAATACATAGCCGAGTAT 915
OY 901 AGCCCTAGCTTGAACGCTGCGGAGATGCGCTGCGCAAAAGCTCCCGCTTTCGTTCA 960
DB 916 GAACAGAAATCTGATGATTAATGCGCTATCAGAGCTGGAATCTGTACCGATCTCACTCG 975
OY 961 TTCCGCTGCCAGGA--TTCCAGCTGCTGCGCTTGTGATTTGTGCGGCTTCATCATAGC 1017
DB 976 TATGCGCCCAAAACAGCCTTGCGCTTATGCTTTTAATCTCGTAACACACGCCCTAT 1035
OY 1018 GATATGTGACGCTGCTGCGGAGATGCGTATGCGCTGCGGCGCGGAGATGCGCT 1077
DB 1036 GATGTGGCAGTTTCTGATTAATTAACGCAATGCTGCGGACGACATCAGTCGCA 1095
OY 1078 CAGCGCTACTGCGAGAAATTAAGCCTAACCAGGACATGCGCGCTTTTGGCCATAT 1137
DB 1096 ATGCCATGTATGCTTATTAACAGCTCCCTGCGATGTGTGCGGCTCCGTCGCAATGAT 1155
OY 1138 AATACAAAGAGTGTGATGCGCTGTGTAATGCCGTTGACCG 1181
DB 1156 AACACCATGAAGAAGTGTGCTGTGAGACGCGCTGCACAG 1199
```

Search completed: April 22, 2003, 10:12:15
Job time : 330.905 secs

QY 301 TAY 111

Db 572 TATGCGGCTCCGGCTGCAACCGGGGATGATATTATTTGTCAGCGTGGCAGAACACAC 631
Oy 361 GCCAACCCTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCAAAGTGGTAATTG 420
Db 632 GCCAACCCTGCTCCCTGCTGATGTCGCCCAACAACCTGAGCCAAAGTGGTAATTG 631
Oy 421 CCGCTTAATGCGAGCGACGCGCGGATGTCGATTTGTTGTCAGAACGATTAATCCCGCT 480
Db 692 CCGCTTAATGCGAGCGACGCGCGGATGTCGATTTGTTGTCAGAACGATTAATCCCGCT 751
Oy 481 AGTCGATTTGCGGCTGTCAGATGTCAGAACGTTACTGCGGTTGCCCGATCTGGCG 540
Db 752 AGTCGATTTGCGGCTGTCAGATGTCAGAACGTTACTGCGGTTGCCCGATCTGGCG 811
Oy 541 CGAGCGATTACCTTTGCTCATTCAGCCGCGATGTCGATGTTGATGTCGTCACGGG 600
Db 812 CGAGCGATTACCTTTGCTCATTCAGCCGCGATGTCGATGTTGATGTCGTCACGGG 871
Oy 601 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATTGATTTGCTTTTTCAGGT 660
Db 872 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATTGATTTGCTTTTTCAGGT 931
Oy 661 CACAACCTGATGCGCGAGAGATGCGGCTGTCGATGTAATCAGAACTGCTGAG 720
Db 932 CACAACCTGATGCGCGAGAGATGCGGCTGTCGATGTAATCAGAACTGCTGAG 991
Oy 721 GCGATGTCGCTTGCTGGCGCGCGCAAAATGTTTCACGAGTCACTTTTGAACGCTTC 780
Db 992 GCGATGTCGCTTGCTGGCGCGCGCAAAATGTTTCACGAGTCACTTTTGAACGCTTC 1051
Oy 781 ACAGCTCAATCTGCGCGTGGAAAGTGGAGCTGGAACGCGCAATGTCGTCGTCATA 840
Db 1052 ACAGCTCAATCTGCGCGTGGAAAGTGGAGCTGGAACGCGCAATGTCGTCGTCATA 1111
Oy 841 GATTAAGCGCGCGCTGGAATGCTGGCAGATTACGATATCAACACGCGCAAGCTGG 900
Db 1112 GATTAAGCGCGCGCTGGAATGCTGGCAGATTACGATATCAACACGCGCAAGCTGG 1171
Oy 901 AGCGCTAGCTTACGACGCTGGCGGAAGATGCGTGGCGAAGCTCCCGCTTTCCTCA 960
Db 1172 AGCGCTAGCTTACGACGCTGGCGGAAGATGCGTGGCGAAGCTCCCGCTTTCCTCA 1231
Oy 961 TTCCGCTGCGAGGATTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 1232 TTCCGCTGCGAGGATTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291
Oy 1021 ATGCTGACGCTGCTGCGGAGATGCGATGTCCTGCGGCGCGCGAGCTGCTGCTGCT 1080
Db 1292 ATGCTGACGCTGCTGCGGAGATGCGATGTCCTGCGGCGCGCGAGCTGCTGCTGCT 1351
Oy 1081 CCGCTACTGCGAGATTAGCGCTAACCGGACACTGCGCGCTTCTTTTGGCCCATATAT 1140
Db 1352 CCGCTACTGCGAGATTAGCGCTAACCGGACACTGCGCGCTTCTTTTGGCCCATATAT 1411
Oy 1141 ACAAGAGTATGTCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1412 ACAAGAGTATGTCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1471
Oy 1201 GATTAA 1206
Db 1472 GATTAA 1477

RESULT 2

US-09-462-645C-5
: Sequence 5, Application US/09462645C
: Patent No. 6436681
: GENERAL INFORMATION:
: APPLICANT: Schroeder, Hartwig
: APPLICANT: Hauer, Bernhard
: TITLE OF INVENTION: The preparation of biotin
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/462,645C

: CURRENT FILING DATE: 2000-01-11
: PRIOR APPLICATION NUMBER: PCT/EP98/04097
: PRIOR FILING DATE: 1998-02-07
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: wordPerfect version 6.1
: SEQ ID NO 5
: LENGTH: 3794
: TYPE: DNA
: ORGANISM: clone pHisBioSI
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 601...1806
US-09-462-645C-5
Query Match 100.0%; Score 1206; DB 4; Length 3794;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ATGACGTTTAAATCCCGCGCAGTTTCCGGCCAGTTTCCGCGACCTACAGATGCGGGC 60
Db 601 ATGACGTTTAAATCCCGCGCAGTTTCCGGCCAGTTTCCGCGACCTACAGATGCGGGC 660
Oy 61 GTCATCTCGACAGCGCGCGGACCGGCTTAAACCTGAAAGCCGTGGTTGAAGCCACCAA 120
Db 661 GTCATCTCGACAGCGCGCGGACCGGCTTAAACCTGAAAGCCGTGGTTGAAGCCACCAA 720
Oy 121 CAGTTTACAGCTGAGCGCGGCAAACTGTCATCCAGACCGCAGTTTCCCGAAGCCCAAGC 180
Db 721 CAGTTTACAGCTGAGCGCGGCAAACTGTCATCCAGACCGCAGTTTCCCGAAGCCCAAGC 780
Oy 181 CTGACGCGCGCTATGACGTGACAGAGAAAGTGGCGCAATTACTGATGACCGGAT 240
Db 781 CTGACGCGCGCTATGACGTGACAGAGAAAGTGGCGCAATTACTGATGACCGGAT 840
Oy 241 GATTAACCTATCTGTCGACGCGCGGACCTGTAATCCATCAACTGATGTCGACATGC 300
Db 841 GATTAACCTATCTGTCGACGCGCGGACCTGTAATCCATCAACTGATGTCGACATGC 900
Oy 301 TATGCGGCTCGCGCTGTCACACCGCGGATGATATTATGTCAGCGTGGCAGAACACAC 360
Db 901 TATGCGGCTCGCGCTGTCACACCGCGGATGATATTATGTCAGCGTGGCAGAACACAC 960
Oy 361 GCGAACCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 961 GCGAACCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Oy 421 CCGCTTAATGCGAGCGACGCTGCGGATGTCGATTTGTTGCGCAGAACGATTAATCCCGCT 480
Db 1021 CCGCTTAATGCGAGCGACGCTGCGGATGTCGATTTGTTGCGCAGAACGATTAATCCCGCT 1080
Oy 481 AGTCGATTTGCGGCTGTCAGATGTCAGAACGTTACTGCGGCTTGGCCGATCTGGCG 540
Db 1081 AGTCGATTTGCGGCTGTCAGATGTCAGAACGTTACTGCGGCTTGGCCGATCTGGCG 1140
Oy 541 CGAGCGATTACCTTTGCTCATTCAGCCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 1141 CGAGCGATTACCTTTGCTCATTCAGCCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Oy 601 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATTGATTTGCTTTTTCAGGT 660
Db 1201 GCAGTGATTTCCCGCGGATGTTCAACCACTGATATTGATTTGCTTTTTCAGGT 1260
Oy 661 CACAACCTGATGCGCGAGAGATGCGGCTGTCGATGTAATCAGAACTGCTGAG 720
Db 1261 CACAACCTGATGCGCGAGAGATGCGGCTGTCGATGTAATCAGAACTGCTGAG 1320
Oy 721 GCGATGTCGCTTGCTGGCGCGCGCAAAATGTTTCACGAGTCACTTTTGAACGCTTC 780
Db 1321 GCGATGTCGCTTGCTGGCGCGCGCAAAATGTTTCACGAGTCACTTTTGAACGCTTC 1380
Oy 781 ACAGCTCAATCTGCGCGTGGAAAGTGGAGCTGGAACGCGCAATGTCGTCGTCATA 840
Db 1381 ACAGCTCAATCTGCGCGTGGAAAGTGGAGCTGGAACGCGCAATGTCGTCGTCATA 1440

QY 841 GGATTAAGCGGCGCTGGAATGGCTGGAGATATACGATATCAACAGCGCGAAAGCTGG 900
DB 1441 GGATTAAGCGGCGCTGGAATGGCTGGAGATATACGATATCAACAGCGCGAAAGCTGG 1500
QY 901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGGCGCTGGCGAAACGTCGCCGCTTTCGTTGA 960
DB 1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGGCGCTGGCGAAACGTCGCCGCTTTCGTTGA 1560
QY 961 TTCGGCTGGCAGGATTCACACCTGCTGGCTTGGATTTGGCTTGCCTTCACTATACGAT 1020
DB 1561 TTCGGCTGGCAGGATTCACACCTGCTGGCTTGGATTTGGCTTGCCTTCACTATACGAT 1620
QY 1021 ATGGTACGCTGGCTGGCGAGATGATGGCTTGGCTGGCGCGCGAGATTCGCTGAG 1080
DB 1621 ATGGTACGCTGGCTGGCGAGATGATGGCTTGGCTGGCGCGCGAGATTCGCTGAG 1680
QY 1081 CCGCTACTGGCAGAAATTAAGCGCTGACCGGACACTGCGCGCTTTTGGCGCATATTAAT 1140
DB 1681 CCGCTACTGGCAGAAATTAAGCGCTGACCGGACACTGCGCGCTTTTGGCGCATATTAAT 1740
QY 1141 ACAAGAGATGATGGATGGCTGGCTGGATGGCTTGCAGCGCGCTGCAATTAATGCTG 1200
DB 1741 ACAAGAGATGATGGATGGCTGGCTGGATGGCTTGCAGCGCGCTGCAATTAATGCTG 1800
QY 1201 GATTAA 1206
DB 1801 GATTAA 1806

RESULT 3
US-09-462-645C-1
Sequence 1, Application US/09462645C
Patent No. 643681
GENERAL INFORMATION:
APPLICANT: Schroeder, Hartwig
APPLICANT: Hauser, Bernhard
TITLE OF INVENTION: The preparation of biotin
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/462,645C
CURRENT FILING DATE: 2000-01-11
PRIORITY APPLICATION NUMBER: PCT/EP98/04097
PRIORITY FILING DATE: 1998-02-07
NUMBER OF SEQ ID NOS: 38
SOFTWARE: WordPerfect version 6.1
SEQ ID NO 1
LENGTH: 1216
TYPE: DNA
ORGANISM: E. coli W3110
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1...11
NAME/KEY: CDS
LOCATION: 12...1216
US-09-462-645C-1

Query Match 99.94% Score 1205 DB 4 Length 1216:
Best Local Similarity 100.0% Pred. No. 0:
Matches 1205; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

DB 192 CTGACCGCGCGCTTAGCAACCTGCACAGAGAAAGTGGCCCAATTCTGTAATCCACCGGAT 251
QY 241 GATTAACATATGCTGGAGCGCGCGCACCTGATTCATGAACATGGTGGCACAATGC 300
DB 252 GATTAACATATGCTGGAGCGCGCGCACCTGATTCATGAACATGGTGGCACAATGC 311
QY 301 TATGCGCTGCCGCTTGCACACCGCGCGATGAGATTAATGTAAGCTGGCAGAACACAC 360
DB 312 TATGCGCGTCCGCTTGCACACCGCGCGATGAGATTAATGTAAGCTGGCAGAACACAC 371
QY 361 GCCAACCTGCTCCCTGCTGATGGTGGCCCAACAACTGAGACCCCAAGTGGTAATGG 420
DB 372 GCCAACCTGCTCCCTGCTGATGGTGGCCCAACAACTGAGACCCCAAGTGGTAATGG 431
QY 421 CCGCTTAATGGCGAGGAGCTGCGCGGATGATGATTTGGTGCAGAACTGATTCCTCCGCT 480
DB 432 CCGCTTAATGGCGAGGAGCTGCGCGGATGATGATTTGGTGCAGAACTGATTCCTCCGCT 491
QY 481 AGTCGATTCCTGGCGTTGGCTCAGATGTCGAACGTTACTGGCGGTTGCCCGATCTGGCG 540
DB 492 AGTCGATTCCTGGCGTTGGCTCAGATGTCGAACGTTACTGGCGGTTGCCCGATCTGGCG 551
QY 541 CGAGCATTAACCTTCTCTCATTCAGCCGGGATGCTGATGGTGTGATGCTGCTCAGGG 600
DB 552 CGAGCATTAACCTTCTCTCATTCAGCCGGGATGCTGATGGTGTGATGCTGCTCAGGG 611
QY 601 GCAGTCGATTTCCCGCGGATGTCAGCAACCTGATTAATGATTTCTATGCTTTTCAGGT 660
DB 612 GCAGTCGATTTCCCGCGGATGTCAGCAACCTGATTAATGATTTCTATGCTTTTCAGGT 671
QY 661 CACAACCTGATGGCGCGACAGCTATCGGCTGCTATGCTTAATCAGAACTGCTGAG 720
DB 672 CACAACCTGATGGCGCGACAGCTATCGGCTGCTATGCTTAATCAGAACTGCTGAG 731
QY 721 GCGATGCTCGCGCTGGCTGGCGCGCAAAATGCTTACGAATGATTTTACGCGCTTC 780
DB 732 GCGATGCTCGCGCTGGCTGGCGCGCAAAATGCTTACGAATGATTTTACGCGCTTC 791
QY 781 AGCAGCTCAATCTGGCGCGTGGAACTGGAAGCTGGAACCGCAATGCTGGTGTGAT 840
DB 792 AGCAGCTCAATCTGGCGCGTGGAACTGGAAGCTGGAACCGCAATGCTGGTGTGAT 851
QY 841 GGATTAAGCGCGCGCTGGAATGGCTGGCAGATTAATCAACAGCGCGAAAGCTGG 900
DB 852 GGATTAAGCGCGCGCTGGAATGGCTGGCAGATTAATCAACAGCGCGAAAGCTGG 911
QY 901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGGCTGGCGGAACGTCGCCGCTTTCGTTCA 960
DB 912 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGGCTGGCGGAACGTCGCCGCTTTCGTTCA 971
QY 961 TTCGGCTGGCAGGATTCACACCTGCTGGCTTGGATTTGGCTTGCCTTCACTATACGAT 1020
DB 972 TTCGGCTGGCAGGATTCACACCTGCTGGCTTGGATTTGGCTTGCCTTCACTATACGAT 1031
QY 1021 ATGGTACGCTGGCTGGCGAGATGATGGCTTGGCTGGCGCGCGAGATTCGCTGAG 1080
DB 1032 ATGGTACGCTGGCTGGCGAGATGATGGCTTGGCTGGCGCGCGAGATTCGCTGAG 1091
QY 1081 CCGCTACTGGCAGAAATTAAGCGCTGAACCGGACACTGCGCGCTTTTGGCGCATATTAAT 1140
DB 1092 CCGCTACTGGCAGAAATTAAGCGCTGAACCGGACACTGCGCGCTTTTGGCGCATATTAAT 1151
QY 1141 ACAAGAGTATGATGGCTGGCTGGATTAATGCTTGAACCGCGCGCTGGAATTAATGGTG 1200
DB 1152 ACAAGAGTATGATGGCTGGCTGGATTAATGCTTGAACCGCGCGCTGGAATTAATGGTG 1211
QY 1201 GATTAA 1205
DB 1212 GATTAA 1216

RESULT 4

OTHER INFORMATION: /note- "N" = adenine or cytosine or
OTHER INFORMATION: guanine or thymine"
US-09-381-862-6

Query Match 12.3% Score 148: DB 4: Length 4973:
Best Local Similarity 48.7% Pred. No. 1e-36:
Matches 559: Conservative 0: Mismatches 571: Indels 19: Gaps 5:

61 GCTATCTGACAGCGCCGACCGCGCTTAAACCTGAAGCCGTGGTGAAGCCACCA 120
1147 GTTATATTTGGATATGCCCCACACACAAACCGCCCTGTCTTGAAGCTTGA 1206
122 CAGTTTACAGTCTGACCGCCGAAACGTCATCGCAGCAGTTGCCCAACCCAC 180
1207 TCTTATTTACAGAGATATATGCTATATGTCACCGAGAGTTCACTTGGCTG 1266
181 CTGACCGCGCTTATGAGCTGACGAGAGAAAGTGCCCAATTAATGATGACCG 240
1267 GCAAGCACAATATGAGGCCAGTGCCGACAGGTGCTGACTTATTCATCTAAT 1326
241 GATAAACTATGCTCTGGACGCGGCGACCACTGAATCCATCAACATGTTGCC 300
1327 AGTAAAGAGTCTCTTTACAGAGAACACACACAGTTGAATGGGTTGCTC- 1383
301 TATGCGCGCTCGCGCTGACCGGCGATGAGATTATGTCAGCGTGGCAGAAC 360
1384 TTTGCGACGACAGCTTTCAGCGCAGAGATGAGGTGATTTGATTAAGAGCA 1443
361 GCCAACCTC-GTCCCTGGCTGATGTCGCCCAACAACTGAGCCAAAGTGAAT 419
1444 GCCAATATCTCTTTGGCAGACGCCGTCAAAAACAGAGAGATTATCTAT 1503
420 GCCGCTTAATGCGCAGCGACTCCGAGTGCATTTGTTGCCAGAACTGATT 479
1504 TTAATTAAG--ATGCGCACTGACATGACATTTGGCAAACTGACGACAA 1560
480 TAGTGGATTTGCGGCTTGGGATGAGATGCAACGTAAGTGGCGGTGCCG 539
1561 AACAGCTTTTGTACCTAGTACATGTCCTCCATGTTCTTGGTCAATATCC 1620
540 GCGAGCAGTACCTTTGCTCATTCAGCCGAGTGGTGTGATGATGATGCTCA 599
1621 ACAAAATGCGCAGCTGGCAGCATGCTAAAGAGCCTACCTGTTGTAAGCG 1680
600 GGCATGCTATTCGCCGCGGATGTCAGCACTGATATGATTTCTATCTTT 659
1681 GCTTCACATTTGGCTATGATGATCAAGACTTGTGATTTCTTTGCTTTT 1740
660 TCACAACTGATGCGCCGACAGATATGCGCGTGTATGTAATAGAACTCT 719
1741 TCATAGATGTTGGGGCCACAGGTTGGGTCTTTAGGCAAGAGACCTTTT 1800
720 GCGGATGTCGCGCTGCTGCGGCGGCAAAATGCTTCAAGAGTGTGACG 779
1801 TCAAGTGAAGCTTGTGAATTTGGCGGAAATGATTTGTTTGAAGAGAG 1860
780 CACGACTAATTTGCGCGCTGAACTGGAAGCTGGAAGCCAAATGTCGCTG 839
1861 CACTTGAAGAAATTTGCTGGAAGTTGAACAGAACACCTCACAATAGCTG 1920
840 AGGATTAAGCGCGC-----GCTGAATGCTGCGCAGATTACGATTAAC 890
1921 TGGGCTAAGCGGAGCATTCTTACCTTCAGAGACTAGGACTGCTGATATCA 1980
891 CGAAAGCTGAGCCCTAGCTTAGCAAGCCTGCGGAGAGATGCGGAAACGTCC 950
1981 TGAAGCAGAACTAATAGCTATGCTTGGCCAAATTAGAGCTATTTAGAG 2040
951 CTTTGTGATTC---CGTGGCAGAGATTCAGCCTGCTGCTTGTGATTTGCT 1007
2041 AATGAGCAACAGCCAGCTAGTGAAGATCTGCTGATTTCTTTAATCTGAG 2100
1008 TCATCATAGCATATGATGAGCGCTGCTGCGGAGTACGTAATGCCCGCGG 1067

DB 2101 GCATCCTCATGACTTGGCAACAGCCTTGACATATGAGCTTTGCAAGAGAGCGCA 2160
OY 1068 GCATGCGCTCAGCCGCTACTGCGAGAAATTAGCGTAACCGGACAGCTGGCGCTTT 1127
DB 2161 CCAGTCGCGCCCAACCTCTTCTTAGTATTTAGCTGTACAGCAACTGTTAGCAAGTT 2220
OY 1128 TCGCGCATATTAACAAAGATGATGTGANTCGCTGTGATGCTGAGCCCGCT 1187
DB 2221 TTAATCTTAAACCAAGCAGATTTGTGACCGCTAGTCAAGCAATTTCTAAAGCAAA 2280
OY 1188 GGAATTATT 1196
DB 2281 GGAGTTTTT 2289

RESULT 7
US-09-221-017B-816/C
Sequence 816, Application US/09221017B
Patent No. 644799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 816:
SEQUENCE CHARACTERISTICS:
LENGTH: 3447 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: circular
TOPOLOGY: circular
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature

```

LOCATION:      1...3447
US-09-221-017B-816

Query Match          7.2% Score 86.4; DB 4; Length 3447;
Best Local Similarity 49.3%; Pred. No. 2.9e-17;
Matches 225; Conservative 0; Mismatches 231; Indels 0; Gaps 0.

OY   60 CCTCTATCTGCACAGCCGCCGCGCACCCGCCCTTAATAACTGAAGCCGTGGTTGAAGCCACCA 119
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   458 CATCATCTTGTGACATGCGCCGACACACAGAAACCATAGCGGTATTAGACCAATTCGA 399
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   120 ACAGTTTATCAGTCTGCAGCCGCGAAGCGTCATCGACAGCCAGTTTGCCGAAGCCCAAG 179
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   398 ACGAGCCTACCGGCACATTCATGCCAATATTCACCCGCCGCTACACCATCTGAGTCAGT 339
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   180 CCTGACCCGCCGCTTATGAGAAGCT*CACGAGAGAAGTGGCGCAATTACTGAATGCACCGGA 239
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   336 GCCTACGGAGGACAACGAAAATGCTCTGCAAGCGGATAGCCCGCTACATCGAGGCGGAAGA 279
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   240 TGAATAAATCTATCGTCTGTGAGCGCGCGGCACCACTGAATCATACATGCTGTGCACAAAT 299
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   278 CCCCAAAGAGGTGCTCTTACCCGCGGACCCACAGAGGCAATCAATCTGTGTGGCTCTTTC 219
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   300 CTATGCGCGTCCGCGCTCGCAACGGGGGATGAGATTATTGTACGCTGTGCAGAACACCA 359
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   218 GTTCGGCCAGCTTTTCGTCCGCTCCGGCGATGAGATTATCATCATGATACCTCGAACACCA 159
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   360 CGCCACACCTCTCCCGCTGCTGATGCTGCCCCAACAACTGGAGCCAAAGTGTGAATTT 419
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   158 TTCGAATATCGTCCCATGGAACATTTTGGCCGACAGTACCGGTGCTCGGCTCAGATGAT 99
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   420 GCCCGTTAATGCGCAGCACTGCGCGATGTCGATTTGTTGCCAGAACTAATTACTCCCG 479
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   98  CCGGTTGACCCAATGGCGGCAACTGGATGGATGGACGATTTCCGCTCTACTGTAACGACG 39
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   480 TAGTCGGATTCGTGGCGTGGGTGAGATGTGCAAGCT 515
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   38  TACCAAAATCGTTTTCGATAGCCCACAGTCACGACATGT 3
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-09-221-017B-801/C
Sequence 801, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998

```

```

PRIOR APPLICATION DATA: PCT/AU98/01023
APPLICATION NUMBER: 10-DEC-1998
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 801:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1...700
US-09-221-0178-801

Query Match 7.18; Score 86; DB 4; Length 700:
Best Local Similarity 49.68; Pred. No. 1.8e-17;
Matches 279; Conservative 0; Mismatches 276; Indels 8; Gaps 2.

OY 627 GCAACTGGATATTGATTTCTATGCTCTTTTCAGGTACACAACGCTGATGGCCCCACAGCTAT 686
Db 686 GGAAGTTTGGAAACGGATTTTCTACGCCCTTCGCGGCATTAAGCTACGGACACAGGAAAT 627
OY 687 CGGCGTGTATGATGTAATACAGAACTGCTGGAGGCGATGTCGGCCGCTGGCGGCGG 746
Db 626 AGCGCTATTGATGGCAAGC--GNAAGTGTTGAAGAAGATCCCCCTACGACGATGGCGG 569
OY 747 CAAATGGTTACGAGAGTGGATTTTGACGCGCTTACGACTCAATCTGCGCGCTGCAACT 806
Db 568 AGAGATGATCGCAAGGTGACATTCGANAAGACCACTTACAGCAATTCCTATTAAGTT 509
OY 807 GGAAGCTGGAACGCCAAATATGCTGCTGTATCATAGATTAAAGCGCGGCTGGATGGCT 866
Db 508 CGAAGCCCGGACGCGGACTATGTGGGCTCCACGGCTTGGCCACAGCTCTGACTACAT 449
OY 867 GGCAGATTACGATTCACACAGCGGGAAGAGCGTACCTTAGCAAGCTGGCGGA 926
Db 448 CGAAGGATTTGGTTTGGAAAGCCATATGCCATGACATGACAGCAAACTCTTGATGGCCAC 389
OY 927 AGATCGCTGGCGGAAGCGCTCCGCGCTTTCGTTCAATTCG-----CTGCAGGATTCAG 980
Db 388 CGAAGCGCTGAGAGCTTTCGCCAATGCGCGTATCATCGGAGGAAGCGGAACAAGAGCGG 329
OY 981 CCTGCTGACCTTTGATTTTGTCTGCGGCTTATCATAGCGATATATGATGACCTCTGGCGGA 1040
Db 328 TGTGCTCTCTTCTTATATAGGGGACATTCATCTATGACCTCGGCAATGCTCTGATCG 269
OY 1041 GTACGCTATTGGCCCGCGCGCGGGGAGATATGGCCTCGCGGCTACTGGCAGAAATTAG 1100
Db 268 GCTGGGTATAGCAATCGGACGAGGACATACATGTCGCCGACCGCTACTGAAAGCATGGG 209
OY 1101 CGTAAACGGGACACTGCGCGCTCTTTTTCGCCCATATATTAACAAGAGTATGGATG 1160
Db 208 ACTCAATGCAACGCGTCTGGCGCTCTTTTCTCTTACATATACGAGGAGGAAGTGCAGCA 149
OY 1161 GCTGGTGAATGCGCTTGACCGCG 1183
Db 148 ATTCTCAAGCGCTTGAAGACGTG 126

```

PCT-US96-05320A-1346
 : Sequence 1346, Application PC/TUS9605320A
 : GENERAL INFORMATION:
 : APPLICANT: Human Genome Sciences
 : APPLICANT: 9410 Key West Avenue
 : APPLICANT: Rockville, MD 20850
 : APPLICANT: United States of America
 : APPLICANT: Johns Hopkins University
 : APPLICANT: 720 Rutland Avenue
 : APPLICANT: Baltimore, MD 21205
 : APPLICANT: United States of America
 : APPLICANT: Mark D. Adams
 : APPLICANT: Owen White
 : APPLICANT: Hamilton O. Smith
 : APPLICANT: J. Craig Ventner
 : TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome.
 : NUMBER OF SEQUENCES: 48
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 : STREET: 1100 New York Avenue, Suite 600
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20003-3934
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 : COMPUTER: HP Vectra 486/33
 : OPERATING SYSTEM: MSDOS version 6.2
 : SOFTWARE: ASCII text
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US96/05320A
 : FILING DATE: April 22, 1996
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/476,102
 : FILING DATE: June 7, 1995
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/487,429
 : FILING DATE: June 7, 1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Eric K. Steife
 : REGISTRATION NUMBER: 36,688
 : REFERENCE/DOCKET NUMBER: 1488, 014PC01
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 371-2600
 : TELEFAX: (202) 371-2540
 : INFORMATION FOR SEQ ID NO: 1346:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 714 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : PCT-US96-05320A-1346

Query Match 6.8%; Score 82.2; DB 5; Length 714;
 Best Local Similarity 48.1%; Pred. No. 2,8e-16;
 Matches 265; Conservative 0; Mismatches 283; Indels 3; Gaps 1:

```

OY 347 TGGCAGAACACCGCCACCTCGTCCCTGCGTATGCTGCGCCCAACAACTGAGCCA 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 TGGTTCATCAAGTGTCTAATTTTGTCACTTGGCATGAAGCGCAAAAAGTGGCGTCAA 175
OY 407 AAGTGTGAATTTGGCGCTTAATGCGAGCGACTGCGGATGTCATTTGTTGCCAGAAC 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 AAATTCGAGTTTACCGCATTTTAAATTAATTTGCTAATTAATGCAAAATGCTTATTCTA 235
OY 467 TGATTACTCCCGTATGATTCGCGTGGGTCAGATGCGAAGTTACTGGCGTT 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 CCGTTTTCGAAAAACAACACTGCTGCGCTGAATTTTGTTCAAATGTCACGTGCAAG 295
OY 527 GCGCGGATCTGGCGGAGCATTAACCTTTGCTCATTCAGCCGG---GATGCTGTGATCG 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 AACACGCGATTAAACGCGCTGATTCAACTTATTAGAAAACATAGTCGCTTGGTTTAA 355
  
```

RESULT 10
 US-09-462-645C-3
 : Sequence 3, Application US/09462645C
 : Patent No. 6436681
 : GENERAL INFORMATION:
 : APPLICANT: Schroeder, Hartwig
 : APPLICANT: Hauser, Bernhard
 : TITLE OF INVENTION: The preparation of biotin
 : FILE REFERENCE:
 : CURRENT APPLICATION NUMBER: US/09/462,645C
 : CURRENT FILING DATE: 2000-01-11
 : PRIOR APPLICATION NUMBER: PCT/EP98/04097
 : PRIOR FILING DATE: 1998-02-07
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: wordperfect version 6.1
 : SEQ ID NO 3
 : LENGTH: 1232
 : TYPE: DNA
 : ORGANISM: E. COLI w3110
 : FEATURE:
 : NAME/KEY: 5'UTR
 : LOCATION: 1...18
 : NAME/KEY: CDS
 : LOCATION: 19...1232
 : US-09-462-645C-3

Query Match 5.1%; Score 61; DB 4; Length 1232;
 Best Local Similarity 46.1%; Pred. No. 2e-09;
 Matches 242; Conservative 0; Mismatches 280; Indels 3; Gaps 1:

```

OY 250 ATGCTTGGACGCGCGCACCTGATTCATCAACATGCTGGCAACATGCTATGCCGT 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 ATGCTTCTTACCTGCTGTGCAACCGAATCTGACAACTGCGGATCAAAAGTGACCAAC 282
OY 310 CCGGCTGTCGAACCGCGGATGATTAATTTTCAGCGGTGGAGAGAACACACGCCAACCTC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 TTTTATATCAAAAAAAGCAACACATCATCAACGAAACGAAACGAAACGAAACGAAACGAAAC 342
OY 370 GTCCCTGCGGATGATGTCGCCCAACAACTGACCAAAAGTGGAATTTGCCCTTAAT 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 GATACCTGCGCTCAGCTGAGACGGGAA---GGTTTGAAGTACCTACCTGAGCACCCAG 399
OY 430 GCGACGCACTGCGGATGTCGATTTGTTGTCAGAACTGATTAATCCCTAGTGGATT 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 CGTAACGCGATTAATGACCTGAAGAACTTAACACGACGATGCTGACGACACCATCTC 459
OY 490 CTGGCGTTGGGTCAAGATGTCAACGCTTACGCGGTTGCCCGATTCGCGCGAGCATTT 549
  
```


Db 1289 GCGCAATGCAAGCGGGGGGTGACAGAGCGGATGCTCCGCC 1333

RESULT 13

US-09-134-001C-788
Sequence 788, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 788

LENGTH: 1164

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-788

Query Match 3.2%: Score 38.2; DB 4; Length 1164;
Best Local Similarity 56.9%: Pred. No. 0.033;

Matches 70: Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Db 583 GTTGATGCTGTCAGGCGGAGTGCATTTCCCGCATGTTCCAGCACTGGATTTGAT 642

Db 541 GTTGATGCTGTTCAAGCAATTGACATTTAGATTTCATTAATTTAAATTTGAT 600

Qy 643 TTCTATGCTTTTTCAGTTCACAACTGTATGCGCCGACAGTATCGCGCTGTATGCT 702

Db 601 ACAATGACTATTTCAGCAGACAAATTTGCTGCTTAAGCTGTGCTACTATTAATA 660

Qy 703 AAA 705

Db 661 AAA 663

RESULT 14

US-08-961-527-6/c
Sequence 6, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland

COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 308-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 20199 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-6

Query Match 3.2%: Score 38; DB 4; Length 20199;
Best Local Similarity 57.6%: Pred. No. 0.16;

Matches 68: Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 583 GTTGATGCTGTCAGGCGGAGTGCATTTCCCGCATGTTCCAGCACTGGATTTGAT 642

Db 13792 GTTGATGCTGTCAGGCGGAGTGCATTTCCCGCATGTTCCAGCACTGGATTTGAT 13733

Qy 643 TTCTATGCTTTTTCAGTTCACAACTGTATGCGCCGACAGTATCGCGCTGTATG 700

Db 13732 TTCTCACTGCTTCTGCGCCACAAATTCATGCTCTTAAGGGAATGCTTCTCTACG 13675

RESULT 15

US-09-403-768-5/c
Sequence 5, Application US/09403768
Patent No. 6444804

GENERAL INFORMATION:

APPLICANT: Lam, Joseph S.
APPLICANT: Burrows, Lori L.

APPLICANT: Walsh, Andrew
APPLICANT: Mawlish, Maurice

TITLE OF INVENTION: No. 6444804el proteins involved in the Synthesis and Assembly

TITLE OF INVENTION: of Core Lipopolysaccharide of Pseudomonas aeruginosa

FILE REFERENCE: 6580-177
CURRENT APPLICATION NUMBER: US/09/403,768

PRIOR FILING DATE: 1999-11-02
PRIOR APPLICATION NUMBER: U.S. 60/045,418

PRIOR FILING DATE: 1997-05-02
PRIOR APPLICATION NUMBER: U.S. 60/046,149

PRIOR FILING DATE: 1997-05-09
NUMBER OF SEQ ID NOS: 8

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 5

LENGTH: 1068
TYPE: DNA

ORGANISM: Pseudomonas aeruginosa rifac (waaC)
FEATURE:

OTHER INFORMATION: Description of Organism: Pseudomonas aeruginosa
US-09-403-768-5

Query Match 2.7%: Score 32.4; DB 4; Length 1068;
Best Local Similarity 50.6%: Pred. No. 2.2;

Matches 78: Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 323 GCGGCGATGATATTATGTCAGCTGGAGCAACAGCCCACTGCTGCTGCA 382

Db 338 CCGGCGAGCGGCGTCTGTCACAGTACGGGTGACGAGCACTCTTACAGACCCCTCG 279

Qy 383 TGGTGGCCCAACAACTGAGCCCAAGTGTGAATGGCCGTTAATGCGCAGGCACTGC 442

Db 278 GCGTGCATCACACAGTATGATGACATTCCTTTCAGCGGCTGTGAAGCGCCGCAATTCG 219

Qy 443 CGGATGTCGATTTTGTGCGCAAACTGATTAATTCG 476

Db 218 CCGTTGCGCAGGCTGTCGCAAGGTTCTTGGCC 185

Search completed: April 22, 2003, 14:27:34

Tue Apr '22 16:31:46 2003

Job time : 134.612 secs

us-09-622-419-3.rni

Page 11

Query Match	Best Local Similarity	Score	DB 10:	Length 1217:
Matches 593: Conservative	51.38%	19.58	0: Mismatches 560: Indels 3: Gaps 2	
OY 51	GAATGGGGGGGCTATTCACAGACGCCCGGACCGGCTTAACCTGAAGCCGCTGTGA	110		
Db 57	GAATCCCTCTTGTCTATCTATCTGCACAGCCGCCGCTACTTGCACAAAGCCATTCAGTATTTGA	116		
OY 111	AGCCACCAACAGATTTCAGTCTGAGCCCGGAAACGTCATCCAGCAGCTTTGCCGA	170		
Db 117	AAAGCTGATGATATTATACCGCGGTACCAATTCTATGTCATCGGAGTGCATACGCT	176		
OY 171	AGCCCAACGCTTACCGCGGCTTATATAAGCTGACACAGAAAGTGGCCATTTACTGAA	230		

Db 177 TGGTACCTTGGCACTGATGATATGAGGCGACGGAAGAAAGTCCGGCTTTCTTGAA 236
OY 231 TGCACCGGATGATAAACTATCTGTGGAGCGCGGACCATGTAATCATCAACATGCT 290
Db 237 TGGCAAGAAAGCGGAGATGTGCTTACTGTCGACAAACGCGCATTAATTAATGAT 296
OY 291 GGCACATGCTATGCGCGCTCGCTGCAACCGGCGGATGATATTTTTCAGCGTGGC 350
Db 297 GGGCAAGATTAAGAAATACCATGTCGGGGAAGGGAAGAAATGTCATTACGGCAAT 356
OY 351 AGAACAACGACGCACTGCTCCCTGGCTGATGTGCGCCCAACAACTGAGCCAAAGT 410
Db 357 GGAACATCATCTCAATATCTATTCATGCGACGACGCTTCCCAAGCGGAAGGGCCGTT 416
OY 411 GGTGAATTTCCGCTTAATGCGGAGCACTGCGGATGTCGATTTGTTGCCAGAACTGAT 470
Db 417 GAAGTATATCCCTTTGCAAGAAAGATGGAGCATTTCTATGCTGATGTGCAACAAACGAT 476
OY 471 TACTCCCGCTAGTCGATTTGGCGTTGGGTCAAGATGTCAACGTTACTGCGGTTGCC 530
Db 477 AACGGAACGAAACAAATTCCTTTCGTTGCTACGTTTCAAAATGTGCTCGGACGCTTAA 536
OY 531 GGAATGCGCGGAGCGATTAATCTTGTCTATGACGCGGATGTCGATGATGATG 590
Db 537 CCCAATTTGCTGAATGCGCAAACTCCGCCATAACATGCGCGGTCAATGCTCATG 596
OY 591 TGTCAAGGGGCAAGTCATTTCCCGCGGATGTTCAACAACATGGAATTTGATATG 650
Db 597 CCGCCAAAGCTGCTCCCATGTCACGCGATGTCACAACTAGACGTACTTTTTCG 656
OY 651 TTTTTCAGTCAAACTGATGCGCGCAAGTATGCGGCTGCTGATGATGATGATGATG 710
Db 657 TTTTTCGCGACAAATGATGCGCTTGGCGATGCGGCTTATACGCAAAAAGC 716
OY 711 ACTGCTGAGGCGATGCGCTTGGCTGGGCGGCGCAAAATGATGTCACGAGTGT 770
Db 717 GTTGCTGAAGCGATGAGCCAAATGTAATTTGGGGGGAATGATGATTTTGTGGTCT 776
OY 771 TGACGCTTACGATCAATCTGCGCGGTGGAACGTAAGGTAAGCGCAAAATGTCG 830
Db 777 GCAGAGATTCACCTTGAAGAGCTCCCATGTAAGTTGAAGGGGGAACGCAATCAT 836
OY 831 TGGGTCATAGATTAAGCGCGGCGTGAATGGCTGCGAGA--TTAGATATCAACGAG 889
Db 837 TGGCGCAATGGGTTGGTCTGCGATTTGTTGTCACGATTTAANGTTTAAAGAAAT 896
OY 890 CCGAAGCTGAGCCCTACCTTAGCAACGCTGCGGAAGATGCGCTGCGCAAACTGCC 949
Db 897 TGAACAGCAGAAAGAACTCGTTGACATTTGCTTGACCGCTCTCGGCATTTCAAGA 956
OY 950 GCTTTGCTGATTCGCTGCCAG--GATTCACGCTGCTGGCTTGTGATTTTCTGGCGT 1007
Db 957 TTTGACGATCTTTGGGCAAAAGACGCTGAGGGGTGATACATTTCCAGCTTGGCGAGT 1016
OY 1008 TCATCATAGGATATGTTGACGCTGCTGGGGAAGTACGATATGCTTGGCGCGCGCA 1067
Db 1017 GCACCCCGGATGACGCTGCTTGTGCTGAAGGAATGCTGCTTGGCGCGCGCA 1076
OY 1068 GCATTCGCTCAGCCGCTTACGAGATTAAGCGTAACCGGACACATGCGCGCTTTT 1127
Db 1077 TCATTTGCAACACGCTTAATGAATGCTTGAACGCTTGAACGCTGCTGCTGCACTA 1136
OY 1128 TGGCCATTAATACAAAGATGATGATGCGCTGCTGTAATGCTGTAACCGCGCT 1187
Db 1137 TTTATGCTACAAATACAAAGATATGATGCTGCTGCAAAAGATTAATCAAAAGAA 1196
OY 1188 GGAATTTTGTGATGAT 1203
Db 1197 GGAATTTTAAAGCAT 1212

RESULT 2
US-09-974-300-4738

Sequence 4738, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4738
LENGTH: 1224
TYPE: DNA
ORGANISM: Bacillus clausii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1224)
OTHER INFORMATION: n - A,T,C or G
US-09-974-300-4738
Query Match 19.5%; Score 234.6; DB 10; Length 1224;
Best Local Similarity 51.3%; Pred. No. 56-69;
Matches 593; Conservative 0; Mismatches 560; Indels 3; Gaps 2;
OY 51 GGAATGCGCGCTATCTGACGACGCGCGGACCGGCTTAACCTGAACCGCTGTTGA 110
Db 57 GAATCTCTTTGCTATCTGACGACGCGCGCATCTTGCAAAACCAATCAACTCATTTGA 116
OY 111 AGCCACCAACAGTATTACAGTCTGAGCGCGGAACGTCATCCGACGATTTGCCGA 170
Db 117 AAGCGTCGATGATTAATACCGCGCTTACAAATTTATATGTCATCGTGGGTGATAGCT 176
OY 171 AGCCCAAGCGCTGACCGCGCTTATGACGCTGACAGAGAGAAAGTGGCGCAATTTACTGA 230
Db 177 TGGTACGCTTGGCAGCATGATGAATGAAGTCGACGCGCAAAAGTCCGGCTTTCTTGA 236
OY 231 TGCACCGGATGATAAACTATGCTGAGCGCGGCGCACCAAGTATCATCAACATGAT 290
Db 237 TGGCGAAGAAAGCGCGAGATGCTTTTACTGTCGACAAACACGCGCATTAATTAATGAT 296
OY 291 GGCACATGCTATGCGCGCTCGCTGCAACCGGCGGATGATATTTTTCAGCGTGGC 350
Db 297 GGGCAAGATTAAGAAATACCATGTCGGGGAAGGGAAGAAATGTCATTACGGCAAT 356
OY 351 AGAACAACGACGCACTGCTCCCTGGCTGATGTGCGCCCAACAACTGAGCCAAAGT 410
Db 357 GGAACATCATCTCAATATCTATTCATGCGACGACGCTTCCCAAGCGGAAGGGCCGTT 416
OY 411 GGTGAATTTCCGCTTAATGCGGAGCACTGCGGATGTCGATTTGTTGCCAGAACTGAT 470
Db 417 GAAGTATATCCCTTTGCAAGAAAGATGGAGCATTTCTATGCTGATGTGCAACAAACGAT 476
OY 471 TACTCCCGCTAGTCGATTTGGCGTTGGGTCAAGATGTCAACGTTACTGCGGTTGCC 530
Db 477 AACGGAACGAAACAAATTCCTTTCGTTGCTACGTTTCAAAATGTGCTCGGACGCTTAA 536
OY 531 GGAATTTGCGCGGAGCAATTAATCTTGTCTCAATTCAGCGGAGATGCTGTAAGTGTGATG 590
Db 537 CCCAATTTGCTGAATGCGCAAACTCCGCCATAACATGCGCGGTATGCTGCTCATG 596
OY 591 TGTCAAGGGGCAAGTCATTTCCCGCGGATGTTCAACAACGTAATGATTTTCAATG 650
Db 597 CCGCCAAAGCTGCTCCCATGTCAGTGAATGTCACAACTTAAGATGATGATTTTGTG 656
OY 651 TTTTTCAGTCAAACTGATGCGCGGACGATGATGCGCTGCTGATGATGATGATGATG 710
Db 657 TTTTTCGCGACAAATGATGCGCTTGGCGATGCGCGCTTATACGCGCAAAAAGC 716

```

Oy 711 ACTGCTGAGCGCATGTCGCCCTGGCGCGCGCAAAATGCTTCAAGAGTGCATT 770
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 GTTGTGTAAGAGATGAGCCCAATTTGATTTGCGGGGAATGATTTGTTGCGTCT 776
Oy 771 TGACGGCTTACAGACTCATATCTGCGCCCTGGAACCTGGAACCTGCAATGTGCG 830
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 777 GCAAGATTTCAGCTTGGAAAGAGCTCCATGCAAGCTTTGAAGGGGAGCCCAATCATTCG 836
Oy 831 TGGCTCATAGGATTAACGGCGCGCTGGATGCTGSCACA-TTACATATCAACAGCG 889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 TGGCCCAATCGGTTGCGTGTGCGCATTTGATTTTCTCACCGATTATGATTTAATGAAT 896
Oy 890 CCGAAGCTGAGCCGCTAGCTTAGCAACGCTGCGGGAAGATGCCCTGGCGAAGCTCCCG 949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 897 TGAACAGACGAAAGAAAGATCGTGTACTATTCTGTTAGCGCTCTCGGCGATTCAAGA 956
Oy 950 GCTTTCGTTCAATTCGCTGCGCAG--GATTCCAGCCTGCTGCGCTTTGATTTTCTGCGCT 1007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 957 TTTGACGATCTTTGGGCGCAAAAGACCGTGCAGCGGTGATCAATTCACGTTGGCGACGT 1016
Oy 1008 TCATCATAGCATATGATGACGCTGCTGGCGAGTACGATATGCGCTGCGGCGCGGCA 1067
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1017 GCACCCCGCAGATGAGGAGCTGCTGTGATGCTGAAGGAATGCTGTGCGGCGCGCA 1076
Oy 1068 GCATTGCGCTGAGCGCTACTGCGAGAAATTAAGCGTAACCGGCACTGCGCGCTCTTT 1127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1077 TCATTGTGACACACCGCTTAATGAATGCTTGACGCTTGCTGCGCACCTCGTCCGACGTA 1136
Oy 1128 TGCCGCATATATACAAAGATGATGCGCTGTGATGCGCTTGACCGCGCGCT 1187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1137 TTATGTATACATACGAAGAAGATATGATGCTGTGCAAAAGGATTAAGTCAAAAGCAA 1196
Oy 1188 GGAATTATTGGTGAT 1203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1197 GGAGTATTTTAGCGAT 1212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 3

US-09-070-927A-290
 : Sequence 290, Application US/09070927A
 : Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch
 Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

```

: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 290:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10397 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 290:
US-09-070-927A-290

```

```

Query Match      16.5%   Score 198.4;   DB 10:   Length 10397;
Best Local Similarity 50.0%;   Pred. No. 2.8e-56;
Matches 583;   Conservative 0;   Mismatches 571;   Indels 12;   Gaps 3;

```

```

Oy 50 AGATGCGGCGCTGATCTGACAGCGCGCGGACCGCGCTTAACCTGAAGCGGTGTG 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 ATGAACCGCTAGTTTATTTAGCAATGCTCGACAAACCAACCAACGCGATTAG 5794
Oy 110 AAGCCACCAAGATTTTACAGTGTGACGCGCGGAAACGTCCTGACGCCAGTTGCCG 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5795 ATGTTTACGACATTTATATGAACCGATTAATGCCAATGTTCAATGCGGTGCACAGCT 5854
Oy 170 AAGCCACCGCTGACCGCGGTTATGAAGCTGACAGAGAAAGTGCGCGAATTACTGA 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 TAGCCGAAGAGACGAGAGACTATGAAGCTCCCGGAAAGGTTCCCAATTTATTC 5914
Oy 230 ATGACCGGATGATTAACATATCTGCTGACGCGCGGCGGACCACTGAATCATCAATG 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5915 ATGCCAAAGAAACAGCAAGACTTTTATTTACAGAGAAACAAACAGTGTGAATTGGA 5974
Oy 290 TGGCACAATGCTATGCGCGCTCGGCTGCAACCGGCGATGATTTTGTACGCTGG 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5975 TTGCCAAAGATTATGTCATTTGTCAGTACAGCGCGGATGAATCTCATATCTTACA 6034
Oy 350 CAGAACCGCGGCAACCTCGCCCTGCGTGTGATGTCGCCCAACCAATGAGGCAAG 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6035 TGAACATCATTTCAACATTTATTCCTGCAACAATTAGTCAACGAGGCGCATTT 6094
Oy 410 TGGTGAATTTGCGCTTATGCGGACGACGCTGCGGATGCGATTTTGGCAGACGTA 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6095 TGAATATATAGCTCACTGAGAGCGCTTTTATGATATGGAAGTGCACGTCACAA 6154
Oy 470 TTACTCCCGTAGTCGATTTCTGCGCTTGGCTGACAGTGTCAACCTTACTGCGGTTGCC 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6155 TTACAGAAAAACGAATAATGTTTGCATGCTCAATGTTCAAAATGCTTACGCGTCATTA 6214
Oy 530 CGGATCTGGCGGACGATTTACCTTGTCTCATTCAGCGGGAATGCTGTGATG 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6215 ATCCAAATGAAGAATTTAACAATTTAGTCTCAAAATGAGCTGTGATGAGAGCG 6274
Oy 590 GTGCTCAGGCGGAGTTCATTTCCCGCGGATGTTACAGCACTGATATTTCTATG 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6275 GCGCCCAAGAGCTTCTCATATGCTGTGAGAGCTTCAAGCCATGCTATTTTATG 6334
Oy 650 CTTTTTCAGTCAACATGTATAGCGCGGAGTATGCGGCTGTATGATGAATAG 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6335 CATTTAGTGTACAAATGTGTGACCAATGCAATGGTGTCTTATATGCAAACTGC 6394
Oy 710 AACTGCTGAGGCGATGTCGCCCTGCGTGCAGCGCGGCAAAATGTTTACCAAGTGA 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6395 ATTTATTTGGAACAAATTTGAACCTTGTGAATTTGGTGAATGATTTTACTTTTCTATC 6454
Oy 770 TTGACGCTTACGACGATCAATCTGCGCGCTGAAACCTGGAAGCTGGAAGCGCAATGTGCG 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6455 TTCAAGAAAGTACTTGGAAAGAGCTTCTTGGAAATTTGAAGCTGCGACACTAATATTTG 6514
Oy 830 CTGCTGTCATAGATTTAAGCGCGGCTGGAATGCGTGGCAGATTTACGATAT-CAA 884
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6515 CCGGGCAATTGCTTAGCTGCGCGCATTTGATTTTAAAGAAATGTTTAAAGAGCGCA 6574

```

Oy 885 CCAGCCGGAACCTGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGC--GCTGGCGA 942
 Db 6575 TTGATCAACATGACAGCGCTTGTTCACATGCTTTTGGCAAGTGGATATGCAAG 6634
 Oy 943 CGTCCGCGGCTTGTTCATTCGCTGAGATTCGA-----GCTGCGGCTTGTGAT 997
 Db 6635 GCTTGAAGTTATGCTGACACAGATCCAAAGACACACAGAGTGTGATTCCTTTAATA 6694
 Oy 998 TTGCTGGCTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1057
 Db 6695 TTGAAGCTTATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6754
 Oy 1058 GGGCGGCGGACATTCGCTGACGCGCTACTGCGAATTAAGGCTTAACCGCACATGC 1117
 Db 6755 GAGCGGCTGACATTCGCGGCAACGCTTATTAACATTTGAGCGCGCACACCGCAC 6814
 Oy 1118 GGGCGCTTTTGGCGCTTAATACAAAGATGATGATGATGATGATGATGATGATGAT 1177
 Db 6815 GGGCAGATTTTATTTATACAAATGACCAAGACGACATGCTTGAAGCATTA 6874
 Oy 1178 ACCGCGCGCTGATTTATGCTGAT 1203
 Db 6875 AACGACAAAGAGTTCCTTCACAT 6900

RESULT 4

US-08-781-986A-81/c
 : Sequence 81, Application US/08781986A
 : Publication No. US20030054436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunesh
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5235
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA

ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII text

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,986A
 FILING DATE:

CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
 NAME: Benson, Bob
 REGISTRATION NUMBER: 30,446
 REFERENCE/DOCKET NUMBER: PB248PP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8512
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ. ID NO: 81:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4280 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

US-08-781-986A-81

Query Match 14.5% Score 175.2: DB 7: Length 4280:
 Best Local Similarity 47.7%: Pred. No. 1.4e-48:
 Matches 550: Conservative 0: Mismatches 593: Indels 9: Gaps 1:

Oy 64 TATCTGACACGCGCGCGCTTAACGCTGAACCGCTGCTTGAAGCACCCACAG 123

Db 3337 TATCTTGATTAACAGCAGCAGTCAAC3CCTGTGCAAGTGTAAATGTTTACAGAT 3278
 Oy 124 TTTACAGCTGAGCGCGGGAAGCTCATGCCAGCACTTTCGGAAGCCACAGCCTG 183
 Db 3277 TACTACAAACGTTATTAATTCAAACGTTCACTGCTGTTCAATGATGATTCGCA 3218
 Oy 184 ACCGGCGCTTATGAGCTGACGAGGAAAGTGGCGCAATTCGTAATGACCCGATGAT 243
 Db 3217 ACTGATGTTATGAAATATCCCTGTAACCTGTCGTTCTTTTATTAATGCAAGTATTT 3158
 Oy 244 AAAACATGCTGAGACGCGCGGACCACTGAATCATCAACATGATGATGATGATGATGAT 303
 Db 3157 GAAAGAAATATTTTAAACAGCGGGAACACTGGCGATTAACCTGTACACATAGCTAT 3098
 Oy 304 GCGCGTCCGCTGTGACACCGCGGATGATTTGTAGCTGTGCAACACACGCGC 363
 Db 3097 GGTGATGCAAAATGTTCAAGAGCGGATGAAATGTTGTACAGAAATGGAACATATGCG 3038
 Oy 364 AACCTGCTCCGCTGCGGTGATGCGCCCAACAACTGAGCCAAAGTGTGAATTCGCG 423
 Db 3037 AATATTGTTCTTGTGCAACAGTTAGCAAAAGCTAAATGCACTTGAATTTATACCA 2978
 Oy 424 CTTAATGCGCAGCAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 483
 Db 2977 ATGACAGCTGAGCGTGAATTAACATGAGATATTAACGAACGATTAATGATAAACA 2918
 Oy 484 CGGATTCGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
 Db 2917 AAGATGCTGCTATGACATATATCTTAATGCTGCGGACAAATTAATATGATTAAC 2858
 Oy 544 GCGATTACCTTTGCTATTCAGCGCGGATGATGATGATGATGATGATGATGATGATGAT 603
 Db 2857 ATTGCAAAATACCTCATACATGCTGCAATTTATGATGATGATGATGATGATGATGAT 2798
 Oy 604 GTGCATTTCCCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
 Db 2797 CCACATATGAACCTGATATGCAAGAAATGAATGCTGATTTATGATGATGATGAT 2738
 Oy 664 AACTGATATGCGCGGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
 Db 2737 AAATGCTTGGACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2678
 Oy 724 ATGTGCGCGGCTGCGGCGGCAAAATGATGATGATGATGATGATGATGATGATGATGAT 783
 Db 2677 ATGGAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2618
 Oy 784 ACTCAATCTGCGCGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCT 843
 Db 2617 TGGCGTGAATTAACCTAAATTTGAAGGGGCTACCTTAATTCCTCAGCAATTTGG 2558
 Oy 844 TTAAGCGCGGCTGGAATGCTGCAATTTACGATATCAACAGCCGCAAGCTGAGC 903
 Db 2557 CTTCGAGAGCTATTCGATTTAGAACCATAGTTTGTGATCAATTCATAATATGAA 2498
 Oy 904 CTAAGCTTGAACAGCTGCGGCGGAAGATGCGCGCAAAAGCTGCGGCTTCTCATTC 963
 Db 2497 CAAGAAATACGATATATCTTATGAGCAAAATGCTGCAATTAAGAAATTTAT 2438
 Oy 964 CGCTGCAGGA-----TTCACCTGCTGCGCTTGTATTTTCTGCGCTTCAATCAT 1014
 Db 2437 GCGCGCGCAAGAGATGCTGCTGCAAGTGAATTAACCTTTATTAAGATGATACCA 2378
 Oy 1015 AGCGATATGATGATGCTGCTGCGGAGTACGATTTGCTGCGCGCGGCGGCAATTCG 1074
 Db 2377 CACGATGCTGCTACAGCCGATGATGAGAGAGTGTAGCGGTGAGACCTGCGCTCAT 2318
 Oy 1075 GCTACGCGCTACTGCGAGATTAAGCGCTAACGCGCAACTGCGCGCTTTTGGCCA 1134
 Db 2317 GCGCAACGTTATTAAGAAATGTTAAATGCTTCAACAGCTGAGAGGATTTTATTA 2258
 Oy 1135 TATATATACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194

Db 2257 TACACAGCAAGAGACGTGATGATTAATGCTTGAACCAAGAGAGTTT 2198
QY 1195 TTGCTGATTA 1206
Db 2197 TTCTCTTATGAA 2186

RESULT 5

US-09-974-300-304
; Sequence 304, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 304
; LENGTH: 727
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-304

Query Match 12.0% Score 145.2; DB 10; Length 727;
Best Local Similarity 51.5% Pred. No. 8,8e-39;

Matches 333; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

QY 60 CCTCTATCTGACAGCGCGCGCGCTTAACCTGAAGCGTGTGAAGCCACCA 119
Db 66 CGTTATTTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
QY 120 ACAGTTTACAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
Db 126 TGAGTTTACAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 185
QY 180 CCTGACCG 239
Db 186 ACCAAGCGTGAATGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
QY 240 TGATAAAGCTATGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
Db 246 CATGCGAAGAAATTTTTCACAAAGCGGTAGCAACGCGCGCGCGCGCG 305
QY 300 CTATGCGCGCTCCGCGTCTGCAACCGCGCGCGCGCGCGCGCGCGCG 359
Db 306 CTATGCG 365
QY 360 CGGCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
Db 366 TGGCAATTAATTTTCCGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 425
QY 420 GCGCGTTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
Db 426 TCGCGTTGAGGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
QY 480 TACTCGGATTTGCGCGTGTGAGATGTCGAAGCTTACTGCGCGCGCG 539
Db 486 TACAAAATGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
QY 540 GCGAGCGATTAAGCTGATTCATTCAGCGCGCGCGCGCGCGCGCGCG 599
Db 546 GGAATGCGCAAAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
QY 600 GCGAGTGCATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 659

Db 606 CACGCCGACATGAATAATGACGTCCAGGNTTGGACTGCTTCTATACGTTTCAACG 665
QY 660 TCACAACATGTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705
Db 666 CCACAAAATGTGTGTCGACCGCGCGCGCGCGCGCGCGCGCGCG 711

RESULT 6

US-09-815-242-3775/c
; Sequence 3775, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Twilck, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA, 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 3775
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-09-815-242-3775

Query Match 11.4% Score 138; DB 10; Length 265;
Best Local Similarity 80.2% Pred. No. 1.5e-36;

Matches 162; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1 ATGACGTTTAAATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 202 ATGACGTTTAAATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143
QY 61 GGTATCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 142 GTTATCTGATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 83
QY 121 CAGTTTACAGTCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 82 CAGTTTATGTTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 23
QY 181 CTGACCGCGCGCGTTATGACGTG 202
Db 22 CTGACCGCGCGCAATATGAAACGG 1

RESULT 7

US-09-738-626-1727
; Sequence 1727, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:


```
Db 209 TGTGACTTTCTTGCGCTTCTTCTCACAAGATGTGTGGCCCTACAGGATTTGATTCCT 268
Oy 696 GATATGTAATATGAGAACTCTGAGAGCCATGTGCCCCCTGGCGGGGGCCCAAAATGCT 755
Db 269 ATACGGTAAATAATACCTTTCTTCTTCTTCATGCGCTCCATTTTGGTGTGGTGAATGAT 328
Oy 756 TCACGAGAGAGTTTGTGAGCGCTTCACGACATCTGCGCGGTGGAACCTGGAACCTGG 815
Db 329 TTCCGATGTATATCTTGTGATCATTTCAACTTATTCGCACTCCACCATATTTGAACCTGG 388
Oy 816 AACGCCAAATGTCTGCTGTGATAGATTA 846
Db 389 AACACCAGCTAGTGGGAGCAATTCGTTTA 419

RESULT 9
US-09-974-300-449
; Sequence 449, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Id Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 449
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-449

Query Match 4.9%; Score 58.8; DB 10; Length 246;
Best Local Similarity 54.1%; Pred. No. 1,1e-09;
Matches 120; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Oy 975 TTCCAGCCGTGCGCTTGTGATTTTCTGCGCTTCATCATAGGATATGTGACCTGCT 1034
Db 23 TCGAGACTCGTACCTTACCTTGAAGACGTTCCGACGAGTGTGGCAACCTGCT 82
Oy 1035 GCGGAGTACGTATTTGCCCTGCGGCGCGGAGCATTTGCGCTCAGCCGCTACTGCGAGA 1094
Db 83 GGATGCAAGAGGGTGGCGGTGAGAGCCGCGCATCTGCGCGCAGCTTTATGAAGATG 142
Oy 1095 ATTAGCGTAAACCGCACACTGCGCGCTTTTGGCCCATATATCAAAAGAGTATGCT 1154
Db 143 GCGAGACGCTGCGCGCGCGCGCAAGACGAGCTTTTATCTGTATAAAGCAAGAGACAT 202
Oy 1155 GATGCGCTGTGAATGCGGTGAGCGCGCGCGGATTAAT 1196
Db 203 CGACAACTGATTAAGCTTTCAGAGACAAAGAGATTTT 244

RESULT 10
US-09-738-626-1726/c
; Sequence 1726, Application US/09738626
; Publication No. US20020197603A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1726
LENGTH: 162
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1726

Query Match 4.6%; Score 55.6; DB 9; Length 162;
Best Local Similarity 59.5%; Pred. No. 1,1e-08;
Matches 94; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Oy 609 TTTCGCGCGGATGTTACGACCACTGGATATGATTTCTATGCTTTTTCAGTCACAACT 668
Db 162 TATGCGAGATTTTCAGACAGCTGATGATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCT 103
Oy 669 GATGAGCCGCGAGCATGATGCGCTGCTGATGTAATCAGAACTGCTGAGCGCATGTC 728
Db 102 GCTGAGACCTGCGAGCGGTGCTGTATCAAAAGTCCCAACTTGTGATGACTGCC 43
Oy 729 CCCCTGCTGGCGCGCGCAAAATGTTTACCAACTGA 766
Db 42 ACCATTTTGTACTGGTGTTCATGATGAAGTTGCA 5

RESULT 11
US-09-938-842A-602
; Sequence 602, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAININ
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,617
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 602
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-602

Query Match 4.2%; Score 50.8; DB 9; Length 1362;
Best Local Similarity 47.7%; Pred. No. 1,4e-06;
Matches 148; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Oy 391 CACAACTGAGCCAAATGTTGATTTCCGCTTAATGCGCAGGATGCGCGATGTC 450
Db 487 CAGCAAGAGAGATTGAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 546
Oy 451 GATTGTGCGAAGTACTGCTCCCGTATGCGATTCGCGCTTGGTGATGATGTCG 510
```

Db 547 GAGATGTTGACAGACGATATTAGGCCACAGACAGGCGTACTTTCTATTATATGCTGTAC 606
OY 511 AACCTTACTGCGGCTTCCCGGATCTGCGCGACGATACCTTTGCTATTCAGCCGG 570
Db 607 AATGAGATGCTGTGTTCAACCTAGTACAGAGATGTAATGATTTGCAAGACATAT 666
OY 571 ATGGTGTATGCTTATGCTGCTCAGAGGGCACTGATTTCCCGCGGATGTTACCAA 630
Db 667 GTTCCGTTTACTACTGATGCTGCTCAAGCTATTGGGAAGTACTGTTGTTAAAG 726
OY 631 CTGATATTGATTTCTATGCTTTTTCAGGTCAACAAGTATGCGCCGACAGATGCGG 690
Db 727 TGGATTTGCTTTGATGCTATGAGTGTCTACAGATCTTATGACCCGAAGGTTGCT 786
OY 691 GTGCTGTATG 700
Db 787 GCTTGTATG 796

RESULT 12

US-08-781-986A-3124
Sequence 3124, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 3124:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-3124

Query Match 4.28: Score 50.2; DB 7: Length 251;
Best Local Similarity 59.4%; Pred. No. 9.5e-07;
Matches 85; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 626 AGCAACTGATATGATTTGATGCTTTTCAGGTCAACAAGTATGCGCCGACAGTA 685
Db 8 AGAATAACTGCTGATTTTATGTTTATGCTCATTAAGTCTTGACACACAGTA 67
OY 686 TCGCGTCTGTATGTAATCAGAACTGCTGAGGCGATGCTCCCTGCTGCGCGCG 745
Db 68 TTGCGTATTATTTGTAACGTAAGTACTACAAAATAAGCAACGATGAGTTCGCTG 127

OY 746 GCAAAATGCTTCACGAAGTCACT 768
Db 128 GCGACATGATGATTTTGTAGT 150

RESULT 13

US-10-081-051-108
Sequence 108, Application US/10081051
Publication No. US20030044422A1
GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Whitmore, William M.
APPLICANT: Kamper, Sondra M.
APPLICANT: Simbi, Bigboy H.
APPLICANT: Ganta, Roman R.
APPLICANT: Moreland, Annie L.
APPLICANT: Mwangli, Duncan M.
APPLICANT: MCGuire, Travis C.
APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Ehrlichia Ruminantium Polypeptides, Antigens, Polynucleotides
FILE REFERENCE: UF-299XC1
CURRENT APPLICATION NUMBER: US/10/081,051
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/269,344
NUMBER OF SEQ ID NOS: 117
SOFTWARE: PatentIn version 3.1
SEQ ID NO 108
LENGTH: 1242
TYPE: DNA
ORGANISM: Ehrlichia ruminantium (formerly Cowdria ruminantium)
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1242)
OTHER INFORMATION: Corresponds to SHQ ID NO:105, nucleotides 2099..3340
OTHER INFORMATION: Hypothetical Iron-sulfur co-factor synthesis
US-10-081-051-108

Query Match 3.3%; Score 39.2; DB 9: Length 1242;
Best Local Similarity 58.6%; Pred. No. 0.011;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

OY 585 TGATGCTCAAGGCGGCTGATTTCCCGGATGTTCAACAAGTATGATTT 644
Db 558 TGATGACGCTCAAGCATTTGTAATAATACCAATAGATGTAACAAATGATATGATTT 617
OY 645 CTATGCTTTTCAAGCTCAACAAGTATGCGCCGACAGATGCGCTGCTATG 700
Db 618 GCTTACTATATGAGACATATAATATATGCTCCAAATGGGAATGGGCAATTATATG 673

RESULT 14

US-10-081-051-105
Sequence 105, Application US/10081051
Publication No. US20030044422A1
GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Whitmore, William M.
APPLICANT: Kamper, Sondra M.
APPLICANT: Simbi, Bigboy H.
APPLICANT: Ganta, Roman R.
APPLICANT: Moreland, Annie L.
APPLICANT: Mwangli, Duncan M.
APPLICANT: MCGuire, Travis C.
APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Ehrlichia Ruminantium Polypeptides, Antigens, Polynucleotides
FILE REFERENCE: UF-299XC1
CURRENT APPLICATION NUMBER: US/10/081,051
CURRENT FILING DATE: 2002-02-20

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:05:34 ; Search time 1966.07 Seconds

(without alignments)
9934.410 Million cell updates/sec

Title: US-09-622-419-3

Perfect score: 1206

Sequence: 1 atgaacgttttaacccgc.....tgaattatctgtgatttaa 1206

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthm:*
3: em_estlm:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_frod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	331	29.1	571	9	AF373211 AF373211
2	333.4	27.6	656	17	CNS03026
3	3320	26.5	932	17	AF095401 AF095401
4	138.4	11.5	755	14	BO518982 EST626397
5	136.8	11.3	725	13	BM407607 EST581922
6	135.4	11.2	645	14	BO578320 WHFO301_E

Result	Score	Query Match	Length	DB ID	Description
7	127.8	10.6	552	10	AV550366
8	126.6	10.5	656	10	AV914366
9	122.6	10.2	713	17	BH242226
10	116.2	9.6	635	17	BH242467
11	115.2	9.6	707	17	BH374662
12	114.8	9.5	591	10	BE511577
13	114.6	9.5	735	17	BH242443
14	114.6	9.5	950	17	BH242437
15	111.6	9.3	669	17	BH242336
16	103.6	8.6	423	17	BH392425
17	103.6	8.6	684	10	AV917435
18	102.6	8.5	596	10	AM459162
19	101	8.4	540	9	AJ470647
20	98	8.1	417	17	BH242524
21	97.4	8.1	542	9	A1773930
22	94.8	7.9	660	14	BO138885
23	92.4	7.7	461	10	AV908881
24	87.8	7.3	666	10	AV822179
25	86.2	7.1	551	10	AV432069
26	85.6	7.1	639	10	AM689739
27	82.6	6.8	530	17	BH242301
28	76.6	6.4	416	10	AV413701
29	76.2	6.3	487	17	BH242276
30	75.8	6.3	803	11	AY104487
31	75.2	6.2	550	10	AW733805
32	75.2	6.2	675	14	BO995735
33	75	6.2	541	14	BM887196
34	74.6	6.2	585	10	AV398793
35	74.6	6.2	777	12	BM645380
36	72.4	6.0	596	10	BM254947
37	71.8	6.0	676	12	BM450659
38	71.2	5.9	349	17	BH242235
39	69.4	5.8	562	12	BF484687
40	68.8	5.7	481	17	AZ049189
41	67	5.6	621	12	BM908875
42	64.4	5.3	358	10	AM966435
43	64.4	5.3	524	17	BM642222
44	63	5.2	258	13	BM442494
45	62.6	5.2	314	17	AZ578055

ALIGNMENTS

RESULT 1
AF373211/c
LOCUS AF373211 571 bp mRNA linear EST 03-MAY-2002
DEFINITION AF373211 Gossypium hirsutum cotyledon Gossypium hirsutum CDNA, mRNA
ACCESSION AF373211
VERSION AF373211.1 GI:20428752
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: Gossypium.

REFERENCE
AUTHORS Li, J., Zhao, G. and Liu, J.
TITLE Isolation and characterization of cDNA fragments induced by salicylic acid in cotton
JOURNAL Unpublished (2002)
COMMENT Contact: Liu Jinyuan
Biological Science and Biotechnology
Tsinghua University
Haidian, Beijing, 100084, China
Email: liujy@mails.tsinghua.edu.cn

FEATURES
source 1.571
location/Qualifiers
/organism="Gossypium hirsutum"
/db_xref="taxon:3635"
/clone_lib="Gossypium hirsutum cotyledon"
/tissue_type="cotyledon"

BASE COUNT 147 a 133 c 150 g 141 t
 ORIGIN /note="Induced by salicylic acid"

Query Match 29.1% Score 351; DB 9; Length 571;
 Best Local Similarity 100.0%; Pred. No. 6e-92;
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 1 ATGAACGTTTAAATCCCGCAGTTTCCGCCAGTTTCCGCCAGTACAGATGCGGC 60
 |||||
 Db 351 ATGAACGTTTAAATCCCGCAGTTTCCGCCAGTTTCCGCCAGTACAGATGCGGC 292

0y 61 GTCTATCTGACAGCGCGCGCGCGCTTAACCTGAAGCCGTGTGAAGCCACCA 120
 |||||
 Db 291 GTCTATCTGACAGCGCGCGCGCGCTTAACCTGAAGCCGTGTGAAGCCACCA 232

0y 121 CAGTTTACAGTCTGAGCGCGCGAGAACTGCACAGTTCGCCAGGCCCAAGC 180
 |||||
 Db 231 CAGTTTACAGTCTGAGCGCGCGAGAACTGCACAGTTCGCCAGGCCCAAGC 172

0y 181 CTGACCGCGCTTATGAAGTGCACAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
 |||||
 Db 171 CTGACCGCGCTTATGAAGTGCACAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 112

0y 241 CATAAACTATGCTGTGACGCGCGCGCACCTGATCATCATCATGCTGCGACATGC 300
 |||||
 Db 111 GATAAACTATGCTGTGACGCGCGCGCACCTGATCATCATCATGCTGCGACATGC 52

0y 301 TATGCGCGCTGCGCTGTGCAACCGCGCGATGATTTGTCAGCGTGCA 351
 |||||
 Db 51 TATGCGCGCTGCGCTGTGCAACCGCGCGATGATTTGTCAGCGTGCA 1

RESULT 2
 CNS01026 656 bp DNA linear GSS 14-JUN-2001
 LOCUS Anopheles gambiae GSS T7 end of clone 25F18 of NotreDame1 library
 DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
 genomic survey sequence.
 ACCESSION AL152975.1 GI:7013894
 VERSION AL152975
 KEYWORDS African malaria mosquito.
 SOURCE Anopheles gambiae
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 656)
 Genoscope.
 AUTHORS Direct Submission
 TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - web : www.genoscope.cns.fr)
 REFERENCE 2 (bases 1 to 656)
 Roth C.W., Brey P.T., Ke Z., Collins F.H. and Weissbach J.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2000) BWHI, Institut Pasteur, 25, rue du Dr.
 Roux, Paris 75015, France
 COMMENT This clone is from an A. gambiae BAC library provided by F.H.
 Collins and sequenced by Genoscope in collaboration with the
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut
 Pasteur.
 FEATURES
 source Location/Qualifiers
 1..656
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="25F18"
 /clone_11b="Notredame1"
 /note="end : 17"
 BASE COUNT 124 a 169 c 218 g 134 t 11 others
 ORIGIN

Query Match 27.6% Score 333.4; DB 17; Length 656;
 Best Local Similarity 73.3%; Pred. No. 9.5e-87;
 Matches 471; Conservative 5; Mismatches 160; Indels 7; Gaps 4;

0y 567 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 626
 |||||
 Db 6 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 65

0y 627 GCAACTGATATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 686
 |||||
 Db 66 GCGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 125

0y 687 GCGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
 |||||
 Db 126 TGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185

0y 747 CAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 806
 |||||
 Db 186 CAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245

0y 807 GGAAGCTGGAAGCGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
 |||||
 Db 246 TGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305

0y 867 GCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924
 |||||
 Db 306 GGAAGGAAAGAAAGCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 365

0y 925 GAAGATGCG 984
 |||||
 Db 366 GAGAGAAAGTAAAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425

0y 985 CTGCGCTTTGATTTGCTGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1044
 |||||
 Db 426 CTGCGCTTTGATTTGCTGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 485

0y 1045 GGTATGCG 1104
 |||||
 Db 486 GGTATGCG 543

0y 1105 ACCGCGACATCG 1164
 |||||
 Db 544 AC--GCGACATCG 601

0y 1165 GTGAATGCGG--TTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1206
 |||||
 Db 602 GTTGGCG 644

RESULT 3
 AF095401/c 932 bp DNA linear GSS 29-AUG-2000
 LOCUS AF095401 Salmonella typhimurium LT2, Lambda DASH II Salmonella
 DEFINITION typhimurium genomic clone 77-T7, DNA sequence.
 ACCESSION AF095401
 VERSION AF095401.1 GI:4322852
 KEYWORDS GSS.
 SOURCE Salmonella typhimurium.
 ORGANISM Salmonella typhimurium
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Salmonella.
 REFERENCE 1 (bases 1 to 932)
 Wong R.M.Y. and McClelland K.
 TITLE End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
 Li-Cor
 JOURNAL Unpublished (1999)
 CONTACT Contact: McClelland M
 COMMENT Molecular Biology
 Sidney Kimmel Cancer Center
 3099 Science Park Road, San Diego, CA 92121, USA
 Email: mclelland@sfsc1.sdsu.edu
 Class: shotgun
 Location/Qualifiers

	source	1..932	/organism="Salmonella typhimurium"
		/strain="LT2"	
		/db_xref="taxon:602"	
		/clone="77-17"	
		/clone_11b="Salmonella typhimurium LT2, Lambda DASH II"	
		/note="Vector: Lambda DASH II; sequenced using Li-Cor sequencer"	
BASE COUNT	220 a	234 c	262 g 215 t 1 others
ORIGIN			
Query Match	26.5%	Score 320;	DB 17; Length 932;
Best Local Similarity	76.5%;	Pred. No. 9.7e-83;	
Matches 406;	Conservative 0;	Mismatches 121;	Indels 4; Gaps 1;
OY	1	ATGAACGTTTTTAAATCCCGCGCAGTTTCGGCCCCAGTTTCCGCACCTACAGCATCGGGC	60
Db	531	ATGAACGCTTTAATCCACGAGTTTGCGGGCAAGTTTCCGGCCTAGCGCATCGGGT	472
OY	61	GTTATCTTCGACAGCGCGCGCACGCCGTAAACCTGAAGCCGTGTTGAAGCCACA	120
Db	471	GTTATCTCGATAGCGCGCGCACGCCGATTAAAGCCACAGCGAGTCATTAGCCACGAC	412
OY	121	CAGTTTACAGTCTAGCGCGCGGAACGTCCTTCGACCGCACTTTGCCGAAGCCCAACGC	180
Db	411	CAGTTTATTGTTTGAGCGCGCGTAACGTTCTACTGACCACTTTGGCCAGCGCGC----	356
OY	181	CTGACCGCGCGGTTTATGAAAGCTGCACAGAAGAGTGGCGCAATTACTGATACCGGAT	240
Db	355	GTGACGCGCGCAATATGAACCGCGCACAGCAAAGCAAGCGCGGCTTTAACCGCGCGAT	296
OY	241	GATAAACTATTCGTGAGCGCGCGCACCACTGAATCCATCAACATGCTGGCACAATGC	300
Db	295	GAAAAAAGTATTCGTGTGACACAGCGCGCACACCAGCAAGCAATGATGATGGCGCAGTGT	236
OY	301	TATGGCGCTCCGCGCTCTGCACCGGGCGCATGAGATTATTGTACGGCTGGCAGAACACCAC	360
Db	235	TACGCCCTCTCTGTCGTCTCGCGCGCGCATGAATATATCGTTAAGCGTGGCCAGCATGAC	176
OY	361	GCAACCTCTGTCGCCCTGATGGTGCAGCAACAACTGAGAGCAAAAGTGTGAATTTG	420
Db	175	GCCAACTCTTGTCCTGCTGATGGTGGCGCACAAACCGCGCCAGCTCTAACACTG	116
OY	421	CGCGTTAATGCGCAGCACTGCCGAGATGCGATTTGTTGCCAGAACTGATTACTCCCGT	480
Db	115	CGCGTTAATGACCGCGGCTTCTGATGTGAGGCGCTGCGCGAATCATCAGTCGCGGC	56
OY	481	AGTCGCAATTCGCGCTTGGGTCAGATGTCGACGTTACTGCGCGGTTGCCCG	531
Db	55	AGCCCGACTCTGGCGCTGGGGCAATGTGCAAGCTCMACCGCGCGCTGCCCG	5
RESULT 4			
LOCUS	BQ518982	755 bp	mRNA linear EST 10-JUN-2002
DEFINITION	EST626397 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STMJN24		
ACCESSION	BQ518982		
VERSION	BQ518982.1		
KEYWORDS	EST.		
ORGANISM	Solanum tuberosum		
SOURCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Asterales; euasterids I; Solanales; Solanaceae; Solanum. 1 (bases 1 to 755) Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Reestepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karavachaya,S.A. Generation of a set of potato cDNA clones for microarray analyses Unpublished (2002) Other_ESTS: EST626398		

[illegible]

RESULT 5
 BM407607 725 bp mRNA linear EST 22-JAN-2002
 LOCUS EST581922 potato roots Solanum tuberosum cDNA clone cPROJ1L21 5'
 DEFINITION end, mRNA sequence.
 ACCESSION BM407607
 VERSION BM407607.1 GI:18259225
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasteridae I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 725)
 van der Hoeven, R., Sun, H., Karamycheva, S.A., Tsai, J., Van Aken, S.,
 Uterback, T., Chienling, A., Bouffri, O., Buell, C.R., Renning, C.,
 Tanksley, S. and Baker, B.
 Generation of ESTs from potato roots
 Title: Research Genetics, Libraries Division
 Contact: Research Genetics, Libraries Division
 Tel: 1-800-711-6195
 Email: cdna@resgen.com
 For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com
 Division tel 1-800-711-6195, email cdna@resgen.com
 Seq primer: 73.
 FEATURES
 source
 1..725
 Location/Qualifiers
 /organism="Solanum tuberosum"
 /cultivar="Kennedec"
 /db_xref="taxon:4113"
 /clone="cPROJ1L21"
 /clone_1lb="potato roots"
 /tissue_type="roots"
 /dev_stage="in vitro grown stem cuttings"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Cornell University, Tanksley Lab;
 sequencing: The Institute for Genomic Research. Roots were
 isolated from in vitro grown stem cuttings on CM medium.
 Roots were isolated two weeks after placing the stem
 cuttings from in vitro grown plants on medium."
 BASE COUNT 184 a 135 c 177 g 229 t
 ORIGIN
 Query Match 11.3% Score 136.8 DB 13 Length 725;
 Best Local Similarity 51.7% Pred. No. 5.7e-29;
 Matches 312; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 629 AACTGATATTGATTTCTATGCTTTTCAGCTCACAAACTGTATGCCCCAGATATCG 688
 DB 371 ATCTAGACACTGATTTCTTCTGTTCTCTCATPAGATGTGGGCTACAGCGCTTG 430
 QY 689 GCGTCTGTATGTAATACAGACTGCTGAGCGCATGTCCCTCGCTGGCCGCGCA 748
 DB 431 GATTCCTGTATGGAAGAGCGCATCTCTTCTGCGATCCCTCTTCTGCTGCTGCG 490
 QY 749 AATGCTTCAAGTGTGATTTGACGGCTTCAGACACTCAATCGCCCGCGAATCG 808
 DB 491 AATGATAGCTGATGTGATTTGATTCATTCATTCGTAACCTCTTCGAGATTG 550
 QY 809 AAGTGCAGCCCAAACTGCTGCTGTCATAGGATTAAGCGCGCCTGGAATGCTG 868
 DB 551 AGGCTGGGACATCCCGCAATGAGAGCAATGAGCAATGAGCAATGATATCTTT 610
 QY 869 CAGA 872
 DB 611 CTGA 614
 RESULT 6
 B0578320 645 bp mRNA linear EST 19-JUN-2002
 LOCUS WHE0301_E11_1212S wheat unstressed seedling shoot cDNA library
 DEFINITION Trilicium aestivum cDNA clone WHE0301_E11_121, mRNA sequence.
 ACCESSION B0578320
 VERSION B0578320.1 GI:21481637
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Trilicium aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triliceae; Trilicium.
 1 (bases 1 to 645)
 Anderson, O.D., Chao, S., Choi, D.N., Close, T.J., Fenton, R.D., Han,
 P.S., Hsia, C.C., Kang, Y., Iazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.
 The structure and function of the expressed portion of the wheat
 genomes - Etisolated shoot cDNA library
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: gandersn@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.
 FEATURES
 source
 1..645
 Location/Qualifiers
 /organism="Trilicium aestivum"
 /cultivar="Chinese 3Spring"
 /db_xref="taxon:4565"
 /clone="WHE0301_E11_121"
 /clone_1lb="wheat unstressed seedling shoot cDNA library"
 /tissue_type="Etisolated shoot"
 /dev_stage="Five day old seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
 and germinated and grown aseptically in the dark at room
 temperature on filter paper with water, mystatin and
 cefotaxime in covered crystallization dishes. Shoots were
 harvested. The tissue, total RNA, and poly(A) RNA were
 prepared. A cDNA library was made, and the cDNA clones
 were in vivo excised to give phagescript phagemids in the
 T3 Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the Olin Anderson lab (all

REFERENCE 1 (bases 1 to 656)
AUTHORS Sato, K., Saitoh, D. and Takeda, K.
TITLE Barley EST sequencing project in NIG and Okayama Univ.
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-1

[illegible]

FEATURES	source
REFERENCE	1.
AUTHORS	Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE	Direct submission of BAC-end sequences from Anopheles gambiae
JOURNAL	unpublished (2001)
COMMENT	Other_GSSS: AG-ND-179H2.TF Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	Other_GSSS: AG-ND-179H2.TF
TITLE	Contact: Brendan J Loftus
JOURNAL	Department of Eukaryotic Genomics
COMMENT	The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	Other_GSSS: AG-ND-179H2.TF
TITLE	Contact: Brendan J Loftus
JOURNAL	Department of Eukaryotic Genomics
COMMENT	The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	Other_GSSS: AG-ND-179H2.TF
TITLE	Contact: Brendan J Loftus
JOURNAL	Department of Eukaryotic Genomics
COMMENT	The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	Other_GSSS: AG-ND-179H2.TF
TITLE	Contact: Brendan J Loftus
JOURNAL	Department of Eukaryotic Genomics
COMMENT	The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	Other_GSSS: AG-ND-179H2.TF
TITLE	Contact: Brendan J Loftus
JOURNAL	Department of Eukaryotic Genomics
COMMENT	The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Class: BAC ends.
ORGANISM	African malaria mosquito.
LOCUS	BH374662 707 bp DNA linear GSS 10-DEC-2001
DEFINITION	AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2, DNA sequence.
ACCESSION	BH374662
VERSION	BH374662.1 GI:17320804
KEYWORDS	GSS.
SOURCE	Anopheles.
ORGANISM	1 (bases 1 to 707) Shetty,J.J., Melek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
REFERENCE	unpublished (2001)
AUTHORS	

BASE COUNT 200 a 158 c 122 g 227 t
 ORIGIN /note="Vector: pECBAC1; Site-1: HindIII"

Query Match 9.6%; Score 115.2; DB 17; Length 707;
 Best Local Similarity 52.0%; Pred. No. 1.3e-22;
 Matches 283; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 328 GATGACATTATTTGTCAGCGTGGCAGACACACGCCACACCTGCTCCCTGGCTGATGTC 387
 11 1111 111 11 111 11 11 111 111 11
 DB 606 GACGACATCATATTTCTTACCTGAGACATCACTTAATATGTTCTTGGCAGATATTG 547
 11 111 111 111 111 111 111 111 111 111
 QY 388 GCCCAACAACCTGAGCAGGAGTGAATGCGGCTTAATGCCAGAGCAGCGCGAT 447
 11 111 111 111 111 111 111 111 111 111
 DB 546 TGTCAAGAACGGGAGCAGAACTTCTGTATATTCATGATGACAGATATTGGCA 487
 11 111 111 111 111 111 111 111 111 111
 QY 448 GTGATTTTGTGCGACAGACTGATTTCTCCCTGATGCTGCTGGCTGATG 507
 11 111 111 111 111 111 111 111 111 111
 DB 486 ATTGATGTTCTGATGATGAGGCTTCTGAGAAACAAACTTCTTCTGTAATCAGCTT 427
 11 111 111 111 111 111 111 111 111 111
 QY 508 TCGAAGCTTACTGCGCGTGGCCGAGTCGCGCGAGCATTAACCTTGTCTCATTGAGCC 567
 11 111 111 111 111 111 111 111 111 111
 DB 426 TCCAAATGCCCTAGCTATTCTTAATCCGATTGATGAATCATCAGAAAGTAGAGCAAA 367
 11 111 111 111 111 111 111 111 111 111
 QY 568 GGGATG---TGGTATGTTGATGTTGCTCAAGGGGCACTGATTTCCCGCGATGTT 624
 11 111 111 111 111 111 111 111 111 111
 DB 366 TCTAATGCTTTTGTATTAATGACGGTGGCAGGCTGACCACTTCAAAATAGATGA 307
 11 111 111 111 111 111 111 111 111 111
 QY 625 CAGCAACTGATATTGATTTCTATGCTTTTTCAGTCAACAACTGATGGCCGACAGT 684
 11 111 111 111 111 111 111 111 111 111
 DB 306 CAGGCAATGATTTGCGATTTCTTCTGCTTCTCAGTCAATTAATGATGACGACCAAGCA 247
 11 111 111 111 111 111 111 111 111 111
 QY 685 ATGCGCTGCTGATGTTAATTCAGAACTGCTGAGCGGATGCGCGCTGCGCGCG 744
 11 111 111 111 111 111 111 111 111 111
 DB 246 ACCGGAATTTTATGCTGAGGAGCGCTTTTAGACACACTTAATCATTCACAGGAGT 187
 11 111 111 111 111 111 111 111 111 111
 QY 745 GCAAAATGCTTACAGAACTGATGTTTACGGGCTTACAGCACTCAATCTCGCGCGTGA 804
 11 111 111 111 111 111 111 111 111 111
 DB 186 GCGAATATGATGACCACTGATGATTTGAGAAACAACTATGACAGACTTCCCTTCA 127
 11 111 111 111 111 111 111 111 111 111
 QY 805 CTGGAAGCTGGAACGCCAATGCTGCTGCTGATGATTAAGCGCGCGCTGGAATG 864
 11 111 111 111 111 111 111 111 111 111
 DB 126 TTGGAAGCGGGAACCAATATTTGAGAGTATATTGGAATGGGAACACGGGATTTT 67
 11 111 111 111 111 111 111 111 111 111
 QY 865 CTGG 868
 111
 DB 66 ATGG 63

RESULT 12

BE511577 591 bp mRNA linear EST 07-AUG-2000
 LOCUS 946062A01.Y1 946 - tassal primordium prepared by Schmidt lab zea
 DEFINITION may cDNA, mRNA sequence.

ACCESSION BE511577
 VERSION BE511577.1 GI:9732825
 KEYWORDS EST

SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 591)
 AUTHORS Walbot V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221

FEATURES
 source
 1..591
 Location/Qualifiers

1..591
 /organism="Zea mays"
 /cultivar="OH43"
 /db_xref="taxon:4577"
 /clone="lib-946 - tassal primordium prepared by Schmidt
 lab"
 /tissue="tassels"
 /dev_stage="just after the transition from vegetative to
 inflorescence development"
 /lab="most-XLOR"
 /note="Organ: tassels; Vector: HybridZAP; Site-1: EcoRI;
 Site-2: XhoI; George Chuck dissected immature tassels
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA
 library in HybridZAP. Sample insert size range was 350 bp
 to 3 kb with a 1 kb average."

BASE COUNT 159 a 110 c 143 g 179 t
 ORIGIN

Query Match 9.5%; Score 114.8; DB 10; Length 591;
 Best Local Similarity 49.8%; Pred. No. 1.5e-22;
 Matches 290; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 269 CCACGTAATCATCAACATGTCGCAACATGCTATGCCGTCGCAACCGCGC 328
 11 1111 11 111 111 111 11 11 11 11 11
 DB 10 CTACTGAACCTATTATATCTGAGCTATTCATGGGACTATCACTGAAGAGAG 69
 11 111 111 111 111 111 111 111 111 111
 QY 329 ATGAGATTATTTGACGCTGGCAGAACACACCCCACTGCTCCCTGGCTGATGTCG 388
 11 111 111 111 111 111 111 111 111 111
 DB 70 ATGAATTTGTTTACAGTTGCAAGCATGATGCTATTGCTTCCATGCGAATGCTTT 129
 11 111 111 111 111 111 111 111 111 111
 QY 389 CCCAACAACCTGAGCAGCAAGTGTGAATTCGCCCTTAATCCGAGGAGCTGCGGATG 448
 11 111 111 111 111 111 11 11 11 11 11
 DB 130 CCCAAGACCTGCTGCCACCTTAATATGTTGATGACTTAAGAAATGTCACAGACA 189
 11 111 111 111 111 111 111 111 111 111
 QY 449 TCGATTGTTGCCAGAACTGATTAATCTCCCGATGCTGCGATTGCGTGGTCAAGATG 508
 11 11 11 11 11 11 11 11 11 11 11
 DB 190 TTGACGAGTTAAAGGTTTCTGTGACAAACAAAGATAGTTGTTGTTCAATGCTCT 249
 11 111 111 111 111 111 111 111 111 111
 QY 509 CGAAGTTTCTGCGGCTGCCCGGATCTGCGCCGACGATTAATCTTGTCTATTACGCC 568
 11 111 111 111 111 111 111 111 111 111
 DB 250 CAATGTTCTAGTTTCATGCTTCCATTCGAGAGATGTAATGATGCTTAACAGAGTTG 309
 11 111 111 111 111 111 111 111 111 111
 QY 569 GGATGCTGTGATGCTGATGCTGCTCAAGGAGGAGATGCTTCCCGCGATGTTCAAC 628
 11 11 11 111 111 11 11 11 11 11 111
 DB 310 GAGCTAAAGCTCTTGTAGATGCTTGCCAAAGTGTTCCTCATATGCGAGTTGATTCAGA 369
 11 111 111 111 111 111 111 111 111 111
 QY 629 AACTGATATTGATTTGATGCTTTTTCAGTTCACAACTGATAGCCCTGACAGTATG 688
 11 111 111 11 11 11 11 11 11 111 111
 DB 370 AACTTGCGCGGACTTTCTTGTGCGATCTTCATTAAGATGTTGGCTTACAGCGCTG 429
 11 111 111 111 111 111 111 111 111 111
 QY 689 GCGTCTGATGCTTAATCAGAACGCTGAGAGCCATGTCGCTGCGCTGGCGCGCGCA 748
 11 111 111 111 111 111 111 111 111 111
 DB 430 GATTTTGCATGTAATTTGAGATTTTGTCTATGAGAGCTTTTATGATGTTGTTG 489
 11 111 111 111 111 111 111 111 111 111
 QY 749 AATGCTTACAGAGTGAATTTTTCAGCGCTTTCACGACTCAATCTCGCGCTGGAATG 808
 11 111 11 11 11 11 11 11 11 11 11
 DB 490 AATGATTCGGAATGATTCGAAGCAATTCACATATGCTCAACCTCCCTTACATTTG 549
 11 111 111 111 111 111 111 111 111 111
 QY 809 AAGCTGGAAGCCCAATGCTGCTGCTCATAGATTAAGCG 850
 11 111 111 11 11 11 11 11 11 11 11
 DB 550 AGCTGGAAGCTCTCAATCGGAGAGATATGAGACTGGAG 591
 11 111 111 111 111 111 111 111 111 111

RESULT 13

BH242443 735 bp DNA linear GSS 13-NOV-2001
 LOCUS BH242443
 DEFINITION AT2FD83TF AT2F Arabidopsis thaliana genomic clone AT2FD83, DNA
 sequence.
 ACCESSION BH242443
 VERSION BH242443.1 GI:16917649

